

4821.1

PATENT

DOCKET 13499 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: KHVOROVA *et al.* Examiner: To be assigned
Serial No.: 10/714333 Group Art Unit: 1646
Filed: November 14, 2003
For: Functional and Hyperfunctional siRNA
Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

April 22, 2005

Mail Stop Petition
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PETITION TO MAKE SPECIAL UNDER 37 C.F.R. § 1.102(d)

Sir:

Pursuant to 37 C.F.R. § 1.102(d) and MPEP 708.02(VIII), Applicants respectfully petition to make special the above-captioned application. A check in the amount of \$130.00 for the fee under 37 C.F.R. § 1.17(h) is enclosed. Applicants submit that all requirements of 37 C.F.R. § 1.102(d) and MPEP 708.02(VIII) have been satisfied, as set forth in the attached Statement in Support of Petition to Make Special.

04/26/2005 MBLAND 00000008 10714333

01 FC:1464

130.00 OP

Certificate of Mailing Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited on the date shown below with the United States Postal Service as first class mail with sufficient postage in an envelope addressed to: Mail Stop Petition, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

(Signature)

TOR SMELAND

(Printed Name of Person Signing Certificate)

22 APRIL 2005

(Date)

Applicant: KHVOROVA *et al.*
Serial No.: 10/714333
Filed: November 14, 2003
Petition to Make Special Under 37 CFR 1.102(d)
April 22, 2005
Page 2 of 2

Attached hereto are the following:

1. a supplemental information disclosure statement;
2. copies of non-patent references cited in the supplemental information disclosure statement;
3. a Statement in Support of Petition to Make Special with attachment Items 1-8;
and
4. a check in the amount of \$130.00 for the petition fee under 37 C.F.R. § 1.17(h).

It is respectfully submitted that all of the requirements of 37 C.F.R. § 1.102(d), 37 C.F.R. § 1.17(h), and MPEP § 708.02(VIII) have been satisfied. Accordingly, it is respectfully requested that this Petition to Make Special be granted and that an accelerated examination of the above-identified application be ordered.

No fee other than the enclosed \$130.00 petition fee under 37 C.F.R. § 1.17(h) is believed to be due in connection with this Petition. If any additional fees are due, or an overpayment has been made, please charge, or credit, our Deposit Account No. 11-0171 for such sum.

If the Examiner has any questions regarding the present Petition or the subject application, the Examiner is cordially invited to contact Applicants' attorney at the telephone number provided below.

Respectfully submitted,



Tor Smeland
Registration No.: 43,131
Attorney for Applicants

Kalow & Springut LLP
Telephone No.: (212) 813-1600



8329.4

PATENT

DOCKET 13499 US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: KHVOROVA *et al.* Examiner: To be assigned
Serial No.: 10/714,333 Group Art Unit: 1646
Filed: November 14, 2003
For: Functional and Hyperfunctional siRNA
Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

April 22, 2005

Mail Stop Petition
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**STATEMENT IN SUPPORT OF
PETITION TO MAKE SPECIAL UNDER 37 C.F.R. § 1.102(d)**

1. Pursuant to the enclosed petition under 37 C.F.R. § 1.102(d) and MPEP 708.02(VIII), Applicants petition to make special the above-captioned application, and provide this Statement in support of the enclosed Petition to Make Special.
2. This is a new application, filed November 14, 2003, which has not undergone any examination by an Examiner.
3. The following items are attached to this Statement:
 - a. A paper copy of the application as filed, along with a copy of a CD-ROM

Certificate of Mailing Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited on the date shown below with the United States Postal Service as first class mail with sufficient postage in an envelope addressed to: Mail Stop Petition, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.



(Signature)

TOR SHELAND

(Printed Name of Person Signing Certificate)

22 APRIL 2005

(Date)

disk containing tables related to nucleic acid listings, which CD-ROM copy was filed with the paper portion of the application on November 14, 2003 as Item 1;

- b. a copy of the Preliminary Amendment amending tables in the paper copy of the application, filed May 20, 2004, as Item 2;
- c. a copy of the Preliminary Amendment directing entry of the sequence listing, filed June 30, 2004, as Item 3;
- d. a copy of the Supplemental Preliminary Amendment canceling claims 9-18, filed by fax on April 21, 2005, as Item 4;
- e. a copy of the International Search Report for PCT/US2003/036787, as Item 5;
- f. a copy of published International Patent Application PCT/US2003/036787, as Item 6;
- g. a copy of the information disclosure statement filed on March 7, 2005, as Item 7; and,
- h. a copy of the information disclosure statement filed on January 31, 2005, as Item 8.

- 4. A supplemental information disclosure statement is filed herewith and enclosed with this mailing, and includes copies of the non-patent references cited in the supplemental information disclosure statement.
- 5. A preexamination search was made on this application. The preexamination search is attached to this Statement as Item 5, which is an International Search Report issued by the U.S. Patent and Trademark Office (USPTO), acting as an International Search Authority, on February 25, 2005 in connection with International Patent Application No. PCT/US2003/036787, filed November 14, 2003, published as WO 2004/045543 A2 on June 3, 2004. A copy of publication No. WO 2004/045543 A2 is attached as Item 6. The disclosure and claims of

International Patent Application No. PCT/US2003/036787 are the same as the disclosure and claims for this application for which special status is requested, with the exception of the attached preliminary amendments made in the instant application.

6. A description of the preexamination search can be found in the International Search Report attached as Item 5, and summarized below:

Classification of Subject Matter

IPC(7): C12Q 1/68; C07H 21/00
US CL: 435/6; 536/24.5

Fields Searched

US: 435/6; 536/24.5

Electronic Data Base Consulted During the International Search

CAPLUS, Medline, BIOSIS, USPATFULL, DERWENT, JPO, EPO

Search Terms: (SIRNA OR RNAI OR DSRNA) and (OPTIMIZATION OR OPTIMIXZE OR OPTIMAL) and ALGORITHM

7. All claims in this application are directed to a single invention.
8. No unity of invention objection has been made to date in PCT/US2003/036787.
9. One copy of each of the references cited in the International Search Report, and a copy of the International Search Report, are transmitted herewith.
10. A discussion of each of the references cited in the International Search Report is provided below. Each of the references cited in the Search Report were

categorized as "A" references. The discussion points out how the claimed subject matter in this application is patentable over each reference.

WO 03/064625 by Woolf and Taylor

11. Applicants submitted International Patent Application Publication No. WO 03/064625 to Woolf and Taylor ("Woolf") by information disclosure statement dated January 31, 2005.
12. The pending claims of the instant application are patentable over Woolf. Woolf—alone or in combination with any other reference herein—does not disclose, teach, or suggest any of the pending claims.
13. Woolf discloses single and double stranded oligonucleotides for inhibiting gene expression. Woolf discloses an oligonucleotide composition having at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene (see for example, page 1, lines 28-30). Woolf discloses that an oligonucleotide composition can have from about 2 oligonucleotides to greater than about 20 oligonucleotides (see, for example, page 26, lines 1-3); that the oligonucleotides can be present at varying concentrations (page 26, lines 11-13); and that they are preferably all double stranded (page 26, lines 16-17). Woolf discloses that the composition inhibits gene expression to an extent that is greater than the level of inhibition of gene expression achieved by any of the individual oligonucleotides of the composition acting alone (page 27, lines 13-15). Regarding selection of double stranded RNAs, Woolf discloses that they should have at least about 60% identity with a target gene (page 18, lines 23-25), and discloses certain alignment criteria for determining identity of a double stranded RNA and a target (page 19).
14. In contrast, claim 1 of the present application recites a method for selecting

siRNA comprising selecting an siRNA of 19-25 nucleoside bases. The method comprises selecting a target gene and measuring the functionality of sequences 19-25 nucleotides in length that are substantially complementary to a stretch of nucleotides in a target sequence, wherein the functionality is dependent upon non-target specific criteria, as recited in claim 1 of the instant application. Among other things, Woolf does not disclose functionality dependent upon non-target specific criteria, as recited in claim 1. Accordingly, Woolf does not disclose, teach, or suggest claim 1. Claims 2-5 depend on claim 1, and thus for at least the reasons that claim 1 is patentable over the cited reference, claims 2-5 are patentable over it as well.

15. Claim 6 of the present application recites a method for developing an siRNA algorithm for selecting siRNA. The method comprises, among other things, determining the amount of improved functionality of an siRNA by the presence or absence of at least one of a number of specified variables. These variables are not disclosed in Woolf. Accordingly, Woolf does not disclose, teach, or suggest claim 6. Claims 7-8 depend on claim 6, and thus, for at least the reasons that claim 6 is patentable over the cited reference, claims 7-8 are patentable over it as well.

16. Claim 19 recites a kit comprising at least two siRNA, wherein the at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein the siRNA are optimized according to recited formulas. Woolf does not disclose, teach, or suggest the recited formulas. Accordingly, Woolf does not disclose, teach, or suggest claim 19.

U.S. Patent Application No. 2002/0150945 A1 by Finney *et al.*

17. The Search Report categorized U.S. Patent Application No. 2002/0150945A1 ("Finney") as an "A" reference, specified pages 27-28, and indicated relevance to

claims 1-19 of this application. Claims 9-18 have been canceled by the Supplemental Preliminary Amendment filed April 22, 2005 (Item 4, attached).

18. The pending claims of the instant application are patentable over Finney. Finney—alone or in combination with any other reference in the Search Report—does not disclose, teach, or suggest any of the pending claims.
19. Finney alleges disclosure of, *inter alia*, dsRNA libraries whereby each dsRNA is capable of reducing the level of expression of a specific gene (see Finney at paragraph [0322], page 27). Finney alleges that more than one “RNAi molecule” may be administered simultaneously or sequentially to a subject or to cells in vitro, asserting that dsRNAs designed to different sequences or regions of a gene can be pooled and administered as one formulation; alternatively, a formulation can comprise dsRNAs that target the mRNA transcripts of different genes (see paragraph [0038], page 28).
20. In contrast, claim 1 of the present application recites a method for selecting siRNA comprising selecting an siRNA of 19-25 nucleoside bases. The method comprises selecting a target gene and measuring the functionality of sequences 19-25 nucleotides in length that are substantially complementary to a stretch of nucleotides in a target sequence, wherein the functionality is dependent upon non-target specific criteria, as recited in claim 1 of the instant application. Among other things, Finney does not disclose functionality dependent upon non-target specific criteria, as recited in claim 1. Accordingly, Finney does not disclose, teach, or suggest claim 1. Claims 2-5 depend on claim 1, and thus for at least the reasons that claim 1 is patentable over the cited reference, claims 2-5 are patentable over it as well.
21. Claim 6 of the present application recites a method for developing an siRNA

algorithm for selecting siRNA. The method comprises, among other things, determining the amount of improved functionality of an siRNA by the presence or absence of at least one of a number of specified variables. These variables are not disclosed in Finney. Accordingly, Finney does not disclose, teach, or suggest claim 6. Claims 7-8 depend on claim 6, and thus, for at least the reasons that claim 6 is patentable over the cited reference, claims 7-8 are patentable over it as well.

22. Claim 19 recites a kit comprising at least two siRNA, wherein the at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein the siRNA are optimized according to recited formulas. Finney does not disclose, teach, or suggest the recited formulas. Accordingly, Finney does not disclose, teach, or suggest claim 19.

Kasif *et al.* (2002) A computational framework for optimal masking in the synthesis of oligonucleotide microarrays. *Nucleic Acids Research*, 30/20:e106

23. The Search Report categorized Kasif *et al.* ("Kasif") as an "A" reference, specified the entire article, and indicated relevance to claims 1-19 of this application. Claims 9-18 have been canceled by the Supplemental Preliminary Amendment filed April 21, 2005 (Item 4, attached).
24. The pending claims of the instant application are patentable over Kasif. Kasif—alone or in combination with any other reference in the Search Report—does not disclose, teach, or suggest any of the pending claims.
25. Kasif discloses a computational formalization of the optimal synthesis strategy for oligonucleotide arrays (see for example, Kasif page 6, first sentence of Conclusions). Kasif is directed to formalizing a synthesis problem for making oligonucleotide arrays, determining precise bounds of the complexity of such a

synthesis, and devising computational solutions (see for example, Kasif at page 1, Abstract). The disclosure focuses on development of an optimal base addition strategy for making an oligo array (see for example, Kasif at page 1, last sentence).

26. In contrast, claim 1 of the present application recites a method for selecting siRNA comprising selecting an siRNA of 19-25 nucleoside bases. The method comprises selecting a target gene and measuring the functionality of sequences 19-25 nucleotides in length that are substantially complementary to a stretch of nucleotides in a target sequence, wherein the functionality is dependent upon non-target specific criteria, as recited in claim 1 of the instant application. Among other things, Kasif does not disclose measuring functionality of any sequences. Accordingly, Kasif does not disclose, teach, or suggest claim 1. Claims 2-5 depend on claim 1, and thus for at least the reasons that claim 1 is patentable over the cited reference, claims 2-5 are patentable over it as well.
27. Claim 6 of the present application recites a method for developing an siRNA algorithm for selecting siRNA comprising, among other things, determining the amount of improved functionality of an siRNA by the presence or absence of at least one of a number of specified variables. These variables are not disclosed in Kasif. Accordingly, Kasif does not disclose, teach, or suggest claim 6. Claims 7 and 8 depend on claim 6, and thus for at least the reasons that claim 6 is patentable over the cited reference, claims 7 and 8 are patentable over it as well.
28. Claim 19 recites a kit comprising at least two siRNA, wherein the at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein the siRNA are optimized according to recited formulas. Kasif does not disclose, teach, or suggest the recited formulas. Accordingly, Kasif does not disclose, teach, or suggest claim 19.

**Amarzguioui *et al.* (2000) Secondary structure prediction and
in vitro accessibility of mRNA as tools in the selection of
target sites for ribozymes. *Nucleic Acids Research*, 28/21, 4113-4124**

29. The Search Report categorized Amarzguioui *et al.* (“Amarzguioui”) as an “A” reference, specified the entire article, and indicated relevance to claims 1-19 of this application. Claims 9-18 have been canceled by Supplemental Preliminary Amendment filed April 21, 2005 (Item 4, attached).
30. The pending claims of the instant application are patentable over Amarzguioui. Amarzguioui —alone or in combination with any other reference in the Search Report—does not disclose, teach, or suggest any of the pending claims.
31. Amarzguioui discloses DNA-armed chemically modified hammerhead ribozymes targeting eight GUC and GUA sites selected by *in vitro* accessibility assays and MFold prediction (a computational method used here to calculate target sequence free energies; see Amarzguioui at page 4121, bottom third of first column). Amarzguioui focuses on target-specific criteria for selecting ribozymes. (see, for example, *Id.*; Amarzguioui at page 4113, Abstract first sentence; Amarzguioui at page 4113, first paragraph of Introduction; Amarzguioui at page 4117, column 2, first full paragraph; and Amarzguioui at page 4118, final paragraph, spanning to the top of page 4119). Amarzguioui discloses that antisense oligodeoxynucleotides can be used to determine accessibility of stretches in mRNA by an *in vitro* assay that measures RNase H-mediated cleavage at sites of hybridization between the antisense oligos and the mRNA (see, for example, Amarzguioui, final paragraph of page 4113 through end of paragraph on page 4114). Amarzguioui also discloses specific structural features of ribozymes that include 8 + 8 nt arms, a stem II structure 2 basepairs long, deoxynucleotides in flanking arms, 2'-O-alkylated residues in the core and stem loop II, a 3' inverted thymidine, a 5' hexanol moiety, and short stretches of phosphorothioate linkages

at one or both ends (see, for example, first full paragraph on page 4114, first column, through end of paragraph in column 2). Amarzguioui investigates correlation between predicted structure of target sites and *in vivo* efficacy of corresponding ribozymes, focusing on target site accessibility parameters that include local free energy of folding of the target sequence, size of single stranded stretches that might function as nucleation sites for duplex formation, and length and stability of stems and helices (see for example, Amarzguioui at page 4121, middle third of first column).

32. In contrast, claim 1 of the present application recites a method for selecting siRNA comprising selecting an siRNA of 19-25 nucleoside bases, comprising selecting a target gene and measuring the functionality of sequences 19-25 nucleotides in length that are substantially complementary to a stretch of nucleotides in a target sequence, wherein the functionality is dependent upon non-target specific criteria. Among other things, Amarzguioui does not disclose selecting an siRNA using a method that comprises measuring functionality, where functionality depends on non-target specific criteria. Accordingly, Amarzguioui does not disclose, teach, or suggest claim 1. Claims 2-5 depend on claim 1, and thus for at least the reasons that claim 1 is patentable over the cited reference, claims 2-5 are patentable over it as well.
33. Claim 6 of the present application recites a method for developing an siRNA algorithm for selecting siRNA comprising, among other things, determining the amount of improved functionality of an siRNA by the presence or absence of at least one of a number of specified variables. These variables are not disclosed in Amarzguioui. Accordingly, Amarzguioui does not disclose, teach, or suggest claim 6. Claims 7 and 8 depend on claim 6, and thus for at least the reasons that claim 6 is patentable over the cited reference, claims 7 and 8 are patentable over it as well.

34. Claim 19 recites a kit comprising at least two siRNA, wherein the at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein the siRNA are optimized according to recited formulas. Amarzguioui does not disclose, teach, or suggest the recited formulas. Accordingly, Amarzguioui does not disclose, teach, or suggest claim 19.
35. Applicants submit that all requirements of 37 C.F.R. 1.102(d), 37 C.F.R. 1.17(h), and MPEP 708.02(VIII) have been satisfied. Accordingly, Applicants request that this application be granted special status.

Respectfully submitted,



Tor Smeland
Registration No.: 43,131
Attorney for Applicants

Kalow & Springut LLP
Telephone No.: (212) 813-1600

ITEM 1

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Khvorova *et al.* Examiner: To be assigned
Serial No.: To be assigned Group Art Unit: To be assigned
Filed: Herewith
For: Functional and Hyperfunctional siRNA

Customer No.:

**23719**

PATENT TRADEMARK OFFICE

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

November 14, 2003

MS Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**TRANSMITTAL OF COMPACT DISK
CONTAINING TABLES AND INCORPORATION BY REFERENCE**

S I R:

Filed herewith are two copies of a compact disk having sequence-related information in Table form. The sequence-related information in Table form on the compact disk is supplied in lieu of a paper submission. Applicants hereby incorporate by reference into the present application the material submitted on the compact disk submitted herewith, in duplicate, on the compact disks labeled DOCKET 13499, PATENT APPLICATION, DISK 1 of 1, COPY 1 of 2; and DOCKET 13499, PATENT APPLICATION, DISK 1 of 1, COPY 2 of 2; which copies are identical, in files entitled Table_12.txt, date of creation June 26, 2003, with a size of 31,045 kb; Table_13.txt, date of creation November 13, 2003, with a size of 78,451 kb; Table_14.txt, date of

Certificate of Express Mailing Under 37 C.F.R. § 1.10

I hereby declare that on the date indicated below, this correspondence is being deposited with the United States Postal Service via Express Mail Label No. EN 035747195 in an envelope addressed to: MS Patent Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA, 22313-1450 on the date shown below.

K Padilla
(Signature)

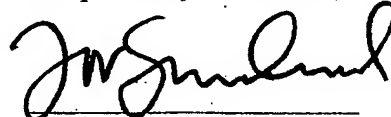
Kim Padilla
(Printed Name of Person Signing Certificate)

11/14/03
(Date)

Applicant: Khvorova, *et al.*
Appl. Ser. No.: To be assigned
Filed: Herewith
Transmittal of Compact Disk Containing Tables
and Incorporation by Reference
November 14, 2003
Page 2 of 2

creation November 13, 2003, with a size of 454 kb; and Table_15.txt date of creation November 13, 2003, with a size of 1,690 kb.

Respectfully submitted,

A handwritten signature in black ink, appearing to read 'Tor Smeland', written over a horizontal line.

Tor Smeland

Registration No.: 43,131

Attorney for Applicants

Kalow & Springut LLP
(212) 813-1600

KALOW & SPRINGUT LLP
488 MADISON AVENUE
NEW YORK, NEW YORK 10022
212-813-1600 212-813-9600
INFO@CREATIVITY-LAW.COM

DOCKET 13499
PATENT APPLICATION
DISK 1 OF 1, COPY 1 OF 2

KALOW & SPRINGUT LLP
488 MADISON AVENUE
NEW YORK, NEW YORK 10022
212-813-1600 212-813-9600
INFO@CREATIVITY-LAW.COM

DOCKET 13499
PATENT APPLICATION
DISK 1 OF 1, COPY 2 OF 2

Functional and Hyperfunctional siRNA**Cross Reference to Related Applications**

5 This application claims the benefit of the filing date of U.S. Provisional Application Serial No. 60/426,137, filed November 14, 2002, entitled "Combinatorial Pooling Approach for siRNA Induced Gene Silencing and Methods for Selecting siRNA," and U.S. Provisional Application Serial No. 60/502,050, filed September 10,
10 2003, entitled "Methods for Selecting siRNA," the entire disclosures of which are hereby incorporated by reference into the present disclosure.

Reference to Tables Submitted in Electronic Form

15 Applicants hereby incorporate by reference the material submitted herewith, in duplicate on the compact disks labeled DOCKET 13499, PATENT APPLICATION, DISK 1 of 1, COPY 1 of 2; and DOCKET 13499, PATENT APPLICATION, DISK 1 of 1, COPY 2 of 2; which copies are identical, in files entitled Table_12.txt, date of creation June 26, 2003, with a size of 31,045 kb; Table_13.txt, date of creation November 13, 2003, with a size of 78,451 kb; Table_14.txt, date of creation
20 November 13, 2003, with a size of 454 kb; and Table_15.txt date of creation November 13, 2003, with a size of 1,690 kb.

Field of Invention

25 The present invention relates to RNA interference ("RNAi").

Background of the Invention

Relatively recently, researchers observed that double stranded RNA ("dsRNA") could be used to inhibit protein expression. This ability to silence a gene has broad potential for treating human diseases, and many researchers and
30 commercial entities are currently investing considerable resources in developing therapies based on this technology.

Double stranded RNA induced gene silencing can occur on at least three different levels: (i) transcription inactivation, which refers to RNA guided DNA or

histone methylation; (ii) siRNA induced mRNA degradation; and (iii) mRNA induced transcriptional attenuation.

It is generally considered that the major mechanism of RNA induced silencing (RNA interference, or RNAi) in mammalian cells is mRNA degradation. Initial attempts to use RNAi in mammalian cells focused on the use of long strands of dsRNA. However, these attempts to induce RNAi met with limited success, due in part to the induction of the interferon response, which results in a general, as opposed to a target-specific, inhibition of protein synthesis. Thus, long dsRNA is not a viable option for RNAi in mammalian systems.

More recently it has been shown that when short (18-30 bp) RNA duplexes are introduced into mammalian cells in culture, sequence-specific inhibition of target mRNA can be realized without inducing an interferon response. Certain of these short dsRNAs, referred to as small inhibitory RNAs ("siRNAs"), can act catalytically at sub-molar concentrations to cleave greater than 95% of the target mRNA in the cell. A description of the mechanisms for siRNA activity, as well as some of its applications are described in Provost *et al.*, *Ribonuclease Activity and RNA Binding of Recombinant Human Dicer*, E.M.B.O. J., 2002 Nov. 1; 21(21): 5864 –5874; Tabara *et al.*, *The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1 and a DexH-box Helicase to Direct RNAi in C. elegans*, Cell 2002, June 28;109(7):861-71; Ketting *et al.*, *Dicer Functions in RNA Interference and in Synthesis of Small RNA Involved in Developmental Timing in C. elegans*; Martinez *et al.*, *Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi*, Cell 2002, Sept. 6; 110(5):563; Hutvagner & Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex*, Science 2002, 297:2056.

From a mechanistic perspective, introduction of long double stranded RNA into plants and invertebrate cells is broken down into siRNA by a Type III endonuclease known as Dicer. Sharp, *RNA interference—2001*, Genes Dev. 2001, 15:485. Dicer, a ribonuclease-III-like enzyme, processes the dsRNA into 19-23 base pair short interfering RNAs with characteristic two base 3' overhangs. Bernstein, Caudy, Hammond, & Hannon, *Role for a bidentate ribonuclease in the initiation step of RNA interference*, Nature 2001, 409:363. The siRNAs are then incorporated into

an RNA-induced silencing complex (RISC) where one or more helicases unwind the siRNA duplex, enabling the complementary antisense strand to guide target recognition. Nykanen, Haley, & Zamore, *ATP requirements and small interfering RNA structure in the RNA interference pathway*, Cell 2001, 107:309. Upon binding to the appropriate target mRNA, one or more endonucleases within the RISC cleaves the target to induce silencing. Elbashir, Lendeckel, & Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs*, Genes Dev 2001, 15:188, **Figure 1**.

The interference effect can be long lasting and may be detectable after many cell divisions. Moreover, RNAi exhibits sequence specificity. Kisielow, M. *et al.* (2002) *Isoform-specific knockdown and expression of adaptor protein ShcA using small interfering RNA*, J. of Biochemistry 363: 1-5. Thus, the RNAi machinery can specifically knock down one type of transcript, while not affecting closely related mRNA. These properties make siRNA a potentially valuable tool for inhibiting gene expression and studying gene function and drug target validation. Moreover, siRNAs are potentially useful as therapeutic agents against: (1) diseases that are caused by over-expression or misexpression of genes; and (2) diseases brought about by expression of genes that contain mutations.

Successful siRNA-dependent gene silencing depends on a number of factors. One of the most contentious issues in RNAi is the question of the necessity of siRNA design, *i.e.*, considering the sequence of the siRNA used. Early work in *C. elegans* and plants circumvented the issue of design by introducing long dsRNA (see, for instance, Fire, A. *et al.* (1998) *Nature* 391:806-811). In this primitive organism, long dsRNA molecules are cleaved into siRNA by Dicer, thus generating a diverse population of duplexes that can potentially cover the entire transcript. While some fraction of these molecules are non-functional (*i.e.* induce little or no silencing) one or more have the potential to be highly functional, thereby silencing the gene of interest and alleviating the need for siRNA design. Unfortunately, due to the interferon response, this same approach is unavailable for mammalian systems. While this effect can be circumvented by bypassing the Dicer cleavage step and directly introducing siRNA, this tactic carries with it the risk that the chosen siRNA sequence may be non-functional or semi-functional.

A number of researches have expressed the view that siRNA design is not a crucial element of RNAi. On the other hand, others in the field have begun to explore the possibility that RNAi can be made more efficient by paying attention to the design of the siRNA. Unfortunately, none of the reported methods have provided a satisfactory scheme for reliably selecting siRNA with acceptable levels of functionality. Accordingly, there is a need to develop rational criteria by which to select siRNA with an acceptable level of functionality, and to identify siRNA that have this improved level of functionality, as well as to identify siRNAs that are hyperfunctional.

Summary of the Invention

The present invention is directed to increasing the efficiency of RNAi, particularly in mammalian systems. Accordingly, the present invention provides kits, siRNAs and methods for increasing siRNA efficacy.

According to one embodiment, the present invention provides a kit for gene silencing, wherein said kit is comprised of a pool of at least two siRNA duplexes, each of which is comprised of a sequence that is complementary to a portion of the sequence of one or more target messenger RNA.

According to a second embodiment, the present invention provides a method for optimizing RNA interference by using one or more siRNAs that are optimized according to a formula (or algorithm) selected from:

Formula I

$$\text{Relative functionality of siRNA} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

Formula II

$$\text{Relative functionality of siRNA} = -(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$$

Formula III

$$\text{Relative functionality of siRNA} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3$$

Formula IV

Relative functionality of siRNA=

$$-GC/2 + (AU_{15-19})/2 - (T_{m20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

5

Formula V

$$\text{Relative functionality of siRNA} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

10 Formula VI

$$\text{Relative functionality of siRNA} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$$

Formula VII

$$\begin{aligned} \text{Relative functionality of siRNA} = & -(GC/2) + (AU_{15-19})/2 - (T_{m20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) \\ 15 \quad & + (A_{19}) * 3 + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2 \end{aligned}$$

wherein in Formulas I – VII:

$T_{m20^{\circ}C} = 1$ if the T_m is greater than $20^{\circ}C$;

20 $A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at positions 15 – 19;

25 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

30 $A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

or

5

Formula VIII Relative functionality of siRNA =

$$(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X; \text{ and }$$

10

Formula IX Relative functionality of siRNA =

$$(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$$

15

wherein

- 20 $A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
25 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
30 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;

$C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;

- $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
5 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;
- 10 $G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
15 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;
- $U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
20 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
25 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;
- 30 $GC_{15-19} =$ the number of G and C bases within positions 15 – 19 of the sense strand or within positions 15 –18 if the sense strand is only 18 base pairs in length;
 $GC_{total} =$ the number of G and C bases in the sense strand;
 $Tm = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs, otherwise its value is 0; and

X = the number of times that the same nucleotide repeats four or more times in a row.

According to a third embodiment, the present invention is directed to a kit comprised of at least one siRNA that contains a sequence that is optimized according to one of the formulas above. Preferably the kit contains at least two optimized siRNA, each of which comprises a duplex, wherein one strand of each duplex comprises at least eighteen contiguous bases that are complementary to a region of a target messenger RNA. For mammalian systems, the siRNA preferably comprises between 18 and 30 nucleotide base pairs.

10

The ability to use the above algorithms, which are not sequence or species specific, allows for the cost-effective selection of optimized siRNAs for specific target sequences. Accordingly, there will be both greater efficiency and reliability in the use of siRNA technologies.

15

According to a fourth embodiment, the present invention provides a method for developing an siRNA algorithm for selecting functional and hyperfunctional siRNAs for a given sequence. The method comprises:

- (a) selecting a set of siRNAs;
- 20 (b) measuring the gene silencing ability of each siRNA from said set;
- (c) determining the relative functionality of each siRNA;
- (d) determining the amount of improved functionality by the presence or absence of at least one variable selected from the group consisting of the total GC content, melting temperature of the siRNA, GC content at positions 15 –19, the presence or absence of a particular nucleotide at a particular position and the number of times that the same nucleotide repeats within a given sequence; and
- 25 (e) developing an algorithm using the information of step (d).

30

According to this embodiment, preferably the set of siRNAs comprises at least 90 siRNAs from at least one gene, more preferably at least 180 siRNAs from at least two different genes, and most preferably at least 270 and 360 siRNAs from at least three and four different genes, respectively. Additionally, in step (d) the determination is made with preferably at least two, more preferably at least three,

even more preferably at least four, and most preferably all of the variables. The resulting algorithm is not target sequence specific.

5 In a fifth embodiment, the present invention provides rationally designed siRNAs identified using the formulas above.

In a sixth embodiment, the present invention is directed to hyperfunctional siRNA.

10 For a better understanding of the present invention together with other and further advantages and embodiments, reference is made to the following description taken in conjunction with the examples, the scope of which is set forth in the appended claims.

15 **Brief Description of the Figures**

Figure 1 shows a model for siRNA-RISC interactions. RISC has the ability to interact with either end of the siRNA or miRNA molecule. Following binding, the duplex is unwound, and the relevant target is identified, cleaved, and released.

20 **Figure 2** is a representation of the functionality of two hundred and seventy siRNA duplexes that were generated to target human cyclophilin, human diazepam-binding inhibitor (DB), and firefly luciferase.

Figure 3a is a representation of the silencing effect of 30 siRNAs in three different cells lines, HEK293, DU145, and Hela. **Figure 3b** shows the frequency of different functional groups (>95% silencing (black), >80% silencing (gray), >50% silencing (dark gray), and <50% silencing (white)) based on GC content. In cases where a given bar is absent from a particular GC percentage, no siRNA were identified for that particular group. **Figure 3c** shows the frequency of different functional groups based on melting temperature (T_m). Again, each group has four different divisions: >95% (black), >80% (gray), >50% (dark gray), and <50% (white) silencing.

25
30

Figure 4 is a representation of a statistical analysis that revealed correlations between silencing and five sequence-related properties of siRNA: (A) an A at position 19 of

the sense strand, (B) an A at position 3 of the sense strand, (C) a U at position 10 of the sense strand, (D) a base other than G at position 13 of the sense strand, and (E) a base other than C at position 19 of the sense strand. All variables were correlated with siRNA silencing of firefly luciferase and human cyclophilin. SiRNAs satisfying the
5 criterion are grouped on the left (Selected) while those that do not, are grouped on the right (Eliminated). Y-axis is “% Silencing of Control.” Each position on the X-axis represents a unique siRNA.

Figures 5 A and 5 B are representations of firefly luciferase and cyclophilin siRNA
10 panels sorted according to functionality and predicted values using Formula VIII. The siRNA found within the circle represent those that have Formula VIII values (SMARTscores™) above zero. SiRNA outside the indicated area have calculated Formula VIII values that are below zero. Y-axis is “Expression (% Control).” Each position on the X-axis represents a unique siRNA.

15 **Figure 6A** is a representation of the average internal stability profile (AISP) derived from 270 siRNAs taken from three separate genes (cyclophilin B, DBI and firefly luciferase). Graphs represent AISP values of highly functional, functional, and non-functional siRNA. **Figure 6B** is a comparison between the AISP of naturally derived
20 GFP siRNA (filled squares) and the AISP of siRNA from cyclophilin B, DBI, and luciferase having >90% silencing properties (no fill) for the antisense strand. “DG” is the symbol for ΔG , free energy.

Figure 7 is a histogram showing the differences in duplex functionality upon
25 introduction of basepair mismatches. The X-axis shows the mismatch introduced in the siRNA and the position it is introduced (e.g., 8C->A reveals that position 8 (which normally has a C) has been changed to an A). The Y-axis is “% Silencing (Normalized to Control).”

30 **Figure 8a** is histogram that shows the effects of 5' sense and antisense strand modification with 2'-O-methylation on functionality. **Figure 8b** is an expression profile showing a comparison of sense strand off-target effects for IGF1R-3 and 2'-O-

methyl IGF1R-3. Sense strand off-targets (lower white box) are not induced when the 5' end of the sense strand is modified with 2'-O-methyl groups (top white box).

5 **Figure 9** shows a graph of SMARTscores™ versus RNAi silencing values for more than 360 siRNA directed against 30 different genes. SiRNA to the right of the vertical bar represent those siRNA that have desirable SMARTscores™.

Figures 10A – E compare the RNAi of five different genes (SEAP, DBI, PLK, Firefly Luciferase, and Renila Luciferase) by varying numbers of randomly selected siRNA and four rationally designed (SMART-selected) siRNA chosen using the algorithm described in Formula VIII. In addition, RNAi induced by a pool of the four SMART-selected siRNA is reported at two different concentrations (100 and 400nM). **10F** is a comparison between a pool of randomly selected EGFR siRNA (Pool 1) and a pool of SMART selected EGFR siRNA (Pool 2). Pool 1, S1—S4 and Pool 2 S1—
15 S4 represent the individual members that made up each respective pool. Note that numbers for random siRNAs represent the position of the 5' end of the sense strand of the duplex. The Y-axis represents the % expression of the control(s). The X-axis is the percent expression of the control.

20 **Figure 11** shows the Western blot results from cells treated with siRNA directed against twelve different genes involved in the clathrin-dependent endocytosis pathway (CHC, DynII, CALM, CLCa, CLCb, Eps15, Eps15R, Rab5a, Rab5b, Rab5c, β 2 subunit of AP-2 and EEA-1). SiRNA were selected using Formula VIII. "Pool" represents a mixture of duplexes 1-4. Total concentration of each siRNA in the pool is 25 nM. Total concentration = $4 \times 25 = 100$ nM.
25

Figure 12 is a representation of the gene silencing capabilities of rationally-selected siRNA directed against ten different genes (human and mouse cyclophilin, C-myc, human lamin A/C, QB (ubiquinol-cytochrome c reductase core protein I), MEK1 and
30 MEK2, ATE1 (arginyl-tRNA protein transferase), GAPDH, and Eg5). The Y-axis is the percent expression of the control. Numbers 1, 2, 3 and 4 represent individual rationally selected siRNA. "Pool" represents a mixture of the four individual siRNA.

Figure 13 is the sequence of the top ten Bcl2 siRNAs as determined by Formula VIII. Sequences are listed 5' to 3'.

Figure 14 is the knockdown by the top ten Bcl2 siRNAs at 100nM concentrations.

- 5 The Y-axis represents the amount of expression relative to the non-specific (ns) and transfection mixture control.

- Figure 15** represents a functional walk where siRNA beginning on every other base pair of a region of the luciferase gene are tested for the ability to silence the luciferase gene. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of each individual siRNA.
- 10

- Figure 16** is a histogram demonstrating the inhibition of target gene expression by pools of 2 and 3 siRNAs duplexes taken from the walk described in Figure 15. The Y-axis represents the percent expression relative to control. The X-axis represents the position of the first siRNA in paired pools, or trios of siRNA. For instance, the first paired pool contains siRNA 1 and 3. The second paired pool contains siRNA 3 and 5. Pool 3 (of paired pools) contains siRNA 5 and 7, and so on.
- 15

- Figure 17** is a histogram demonstrating the inhibition of target gene expression by pools of 4 and 5 siRNA duplexes. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of the first siRNA in each pool.
- 20

- Figure 18** is a histogram demonstrating the inhibition of target gene expression by siRNAs that are ten and twenty basepairs apart. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of the first siRNA in each pool.
- 25

- Figure 19** shows that pools of siRNAs (dark gray bar) work as well (or better) than the best siRNA in the pool (light gray bar). The Y-axis represents the percent expression relative to a control. The X axis represents the position of the first siRNA in each pool.
- 30

Figure 20 shows that the combination of several semifunctional siRNAs (dark gray) result in a significant improvement of gene expression inhibition over individual (semi-functional; light gray) siRNA. The Y-axis represents the percent expression relative to a control.

5

Figure 21 shows both pools (Library, Lib) and individual siRNAs in inhibition of gene expression of Beta-Galactosidase, Renilla Luciferase and SEAP (alkaline phosphatase). Numbers on the X-axis indicate the position of the 5'-most nucleotide of the sense strand of the duplex. The Y-axis represents the percent expression of each gene relative to a control. Libraries contain siRNAs that begin at the following nucleotides: Seap: Lib 1: 206, 766, 812, 923, Lib 2: 1117, 1280, 1300, 1487, Lib 3: 206, 766, 812, 923, 1117, 1280, 1300, 1487; Lib 4: 206, 812, 1117, 1300, Lib 5: 766, 923, 1280, 1487, Lib 6: 206, 1487; Bgal: Lib 1: 979, 1339, 2029, 2590, Lib 2: 1087, 1783, 2399, 3257, Lib 3: 979, 1783, 2590, 3257, Lib 4: 979, 1087, 1339, 1783, 2029, 2399, 2590, 3257, Lib 5: 979, 1087, 1339, 1783, Lib 6: 2029, 2399, 2590, 3257; Renilla: Lib 1: 174, 300, 432, 568, Lib 2: 592, 633, 729, 867, Lib 3: 174, 300, 432, 568, 592, 633, 729, 867, Lib 4: 174, 432, 592, 729, Lib 5: 300, 568, 633, 867, Lib 6: 592, 568.

10

15

20

Figure 22 shows the results of an EGFR and TfnR internalization assay when single gene knockdowns are performed. The Y-axis represents percent internalization relative to control.

25

Figure 23 shows the results of an EGFR and TfnR internalization assay when multiple genes are knocked down (e.g. Rab5a, b, c). The Y-axis represents the percent internalization relative to control.

30

Figure 24 shows the simultaneous knockdown of four different genes. SiRNAs directed against G6PD, GAPDH, PLK, and UBQ were simultaneously introduced into cells. Twenty-four hours later, cultures were harvested and assayed for mRNA target levels for each of the four genes. A comparison is made between cells transfected with individual siRNAs vs. a pool of siRNAs directed against all four genes.

Figure 25 shows the functionality of ten siRNAs at 0.3nM concentrations.

Detailed Description

Definitions

5 Unless stated otherwise, the following terms and phrases have the meanings provided below:

siRNA

10 The term “siRNA” refers to small inhibitory RNA duplexes that induce the RNA interference (RNAi) pathway. These molecules can vary in length (generally between 18-30 basepairS) and contain varying degrees of complementarity to their target mRNA in the antisense strand. Some, but not all, siRNA have unpaired overhanging bases on the 5’ or 3’ end of the sense strand and/or the antisense strand. The term “siRNA” includes duplexes of two separate strands, as well as single strands
15 that can form hairpin structures comprising a duplex region.

SiRNA may be divided into five (5) groups (non-functional, semi-functional, functional, highly functional, and hyper-functional) based on the level or degree of silencing that they induce in cultured cell lines. As used herein, these definitions are
20 based on a set of conditions where the siRNA is transfected into said cell line at a concentration of 100nM and the level of silencing is tested at a time of roughly 24 hours after transfection, and not exceeding 72 hours after transfection. In this context, “non-functional siRNA” are defined as those siRNA that induce less than 50% (<50%) target silencing. “Semi-functional siRNA” induce 50-79% target silencing.
25 “Functional siRNA” are molecules that induce 80-95% gene silencing. “Highly-functional siRNA” are molecules that induce greater than 95% gene silencing. “Hyperfunctional siRNA” are a special class of molecules. For purposes of this document, hyperfunctional siRNA are defined as those molecules that: (1) induce greater than 95% silencing of a specific target when they are transfected at
30 subnanomolar concentrations (*i.e.*, less than one nanomolar); and/or (2) induce functional (or better) levels of silencing for greater than 96 hours. These relative functionalities (though not intended to be absolutes) may be used to compare siRNAs to a particular target for applications such as functional genomics, target identification and therapeutics.

miRNA

The term “miRNA” refers to microRNA.

5 Gene silencing

The phrase “gene silencing” refers to a process by which the expression of a specific gene product is lessened or attenuated. Gene silencing can take place by a variety of pathways. Unless specified otherwise, as used herein, gene silencing refers to decreases in gene product expression that results from RNA interference (RNAi), a defined, though partially characterized pathway whereby small inhibitory RNA (siRNA) act in concert with host proteins (*e.g.* the RNA induced silencing complex, RISC) to degrade messenger RNA (mRNA) in a sequence-dependent fashion. The level of gene silencing can be measured by a variety of means, including, but not limited to, measurement of transcript levels by Northern Blot Analysis, B-DNA techniques, transcription-sensitive reporter constructs, expression profiling (*e.g.* DNA chips), and related technologies. Alternatively, the level of silencing can be measured by assessing the level of the protein encoded by a specific gene. This can be accomplished by performing a number of studies including Western Analysis, measuring the levels of expression of a reporter protein that has *e.g.* fluorescent properties (*e.g.* GFP) or enzymatic activity (*e.g.* alkaline phosphatases), or several other procedures.

Transfection

The term “transfection” refers to a process by which agents are introduced into a cell. The list of agents that can be transfected is large and includes, but is not limited to, siRNA, sense and/or anti-sense sequences, DNA encoding one or more genes and organized into an expression plasmid, proteins, protein fragments, and more. There are multiple methods for transfecting agents into a cell including, but not limited to, electroporation, calcium phosphate-based transfections, DEAE-dextran-based transfections, lipid-based transfections, molecular conjugate-based transfections (*e.g.* polylysine-DNA conjugates), microinjection and others.

Target

The term “target” is used in a variety of different forms throughout this document and is defined by the context in which it is used. “Target mRNA” refers to a messenger RNA to which a given siRNA can be directed against. “Target sequence” and “target site” refer to a sequence within the mRNA to which the sense strand of an siRNA shows varying degrees of homology and the antisense strand exhibits varying degrees of complementarity. The term “siRNA target” can refer to the gene, mRNA, or protein against which an siRNA is directed. Similarly “target silencing” can refer to the state of a gene, or the corresponding mRNA or protein.

Off-target silencing and Off-target interference

The phrases “off-target silencing” and “off-target interference” are defined as degradation of mRNA other than the intended target mRNA due to overlapping and/or partial homology with secondary mRNA messages.

SMARTscore™

The term “SMARTscore™” refers to a number determined by applying any of the Formulas I - Formula IX to a given siRNA sequence. The term “SMART-selected” or “rationally selected” or “rational selection” refers to siRNA that have been selected on the basis of their SMARTscores™.

Complementary

The term “complementary” refers to the ability of polynucleotides to form base pairs with one another. Base pairs are typically formed by hydrogen bonds between nucleotide units in antiparallel polynucleotide strands. Complementary polynucleotide strands can base pair in the Watson-Crick manner (*e.g.*, A to T, A to U, C to G), or in any other manner that allows for the formation of duplexes. As persons skilled in the art are aware, when using RNA as opposed to DNA, uracil rather than thymine is the base that is considered to be complementary to adenosine. However, when a U is denoted in the context of the present invention, the ability to substitute a T is implied, unless otherwise stated.

Perfect complementarity or 100% complementarity refers to the situation in which each nucleotide unit of one polynucleotide strand can hydrogen bond with a

nucleotide unit of a second polynucleotide strand. Less than perfect complementarity refers to the situation in which some, but not all, nucleotide units of two strands can hydrogen bond with each other. For example, for two 20-mers, if only two base pairs on each strand can hydrogen bond with each other, the polynucleotide strands exhibit 10% complementarity. In the same example, if 18 base pairs on each strand can hydrogen bond with each other, the polynucleotide strands exhibit 90% complementarity. "Substantial complementarity" refers to polynucleotide strands exhibiting 79% or greater complementarity, excluding regions of the polynucleotide strands, such as overhangs, that are selected so as to be noncomplementary. ("Substantial similarity" refers to polynucleotide strands exhibiting 79% or greater similarity, excluding regions of the polynucleotide strands, such as overhangs, that are selected so as not to be similar.) Thus, for example, two polynucleotides of 29 nucleotide units each, wherein each comprises a di-dT at the 3' terminus such that the duplex region spans 27 bases, and wherein 26 of the 27 bases of the duplex region on each strand are complementary, are substantially complementary since they are 96.3% complementary when excluding the di-dT overhangs.

Deoxynucleotide

The term "deoxynucleotide" refers to a nucleotide or polynucleotide lacking a hydroxyl group (OH group) at the 2' and/or 3' position of a sugar moiety. Instead, it has a hydrogen bonded to the 2' and/or 3' carbon. Within an RNA molecule that comprises one or more deoxynucleotides, "deoxynucleotide" refers to the lack of an OH group at the 2' position of the sugar moiety, having instead a hydrogen bonded directly to the 2' carbon.

Deoxyribonucleotide

The terms "deoxyribonucleotide" and "DNA" refer to a nucleotide or polynucleotide comprising at least one sugar moiety that has an H, rather than an OH, at its 2' and/or 3' position.

Substantially Similar

The phrase "substantially similar" refers to a similarity of at least 90% with respect to the identity of the bases of the sequence.

Duplex Region

The phrase “duplex region” refers to the region in two complementary or substantially complementary polynucleotides that form base pairs with one another, either by Watson-Crick base pairing or any other manner that allows for a stabilized duplex between polynucleotide strands that are complementary or substantially complementary. For example, a polynucleotide strand having 21 nucleotide units can base pair with another polynucleotide of 21 nucleotide units, yet only 19 bases on each strand are complementary or substantially complementary, such that the “duplex region” has 19 base pairs. The remaining bases may, for example, exist as 5’ and 3’ overhangs. Further, within the duplex region, 100% complementarity is not required; substantial complementarity is allowable within a duplex region. Substantial complementarity refers to 79% or greater complementarity. For example, a mismatch in a duplex region consisting of 19 base pairs results in 94.7% complementarity, rendering the duplex region substantially complementary.

Nucleotide

The term “nucleotide” refers to a ribonucleotide or a deoxyribonucleotide or modified form thereof, as well as an analog thereof. Nucleotides include species that comprise purines, *e.g.*, adenine, hypoxanthine, guanine, and their derivatives and analogs, as well as pyrimidines, *e.g.*, cytosine, uracil, thymine, and their derivatives and analogs.

Nucleotide analogs include nucleotides having modifications in the chemical structure of the base, sugar and/or phosphate, including, but not limited to, 5-position pyrimidine modifications, 8-position purine modifications, modifications at cytosine exocyclic amines, and substitution of 5-bromo-uracil; and 2’-position sugar modifications, including but not limited to, sugar-modified ribonucleotides in which the 2’-OH is replaced by a group such as an H, OR, R, halo, SH, SR, NH₂, NHR, NR₂, or CN, wherein R is an alkyl moiety. Nucleotide analogs are also meant to include nucleotides with bases such as inosine, queuosine, xanthine, sugars such as 2’-methyl ribose, non-natural phosphodiester linkages such as methylphosphonates, phosphorothioates and peptides.

Modified bases refer to nucleotide bases such as, for example, adenine, guanine, cytosine, thymine, uracil, xanthine, inosine, and queuosine that have been modified by the replacement or addition of one or more atoms or groups. Some examples of types of modifications that can comprise nucleotides that are modified with respect to the base moieties include but are not limited to, alkylated, halogenated, thiolated, aminated, amidated, or acetylated bases, individually or in combination. More specific examples include, for example, 5-propynyluridine, 5-propynylcytidine, 6-methyladenine, 6-methylguanine, N,N,-dimethyladenine, 2-propyladenine, 2-propylguanine, 2-aminoadenine, 1-methylinosine, 3-methyluridine, 5-methylcytidine, 5-methyluridine and other nucleotides having a modification at the 5 position, 5-(2-amino)propyl uridine, 5-halocytidine, 5-halouridine, 4-acetylcytidine, 1-methyladenosine, 2-methyladenosine, 3-methylcytidine, 6-methyluridine, 2-methylguanosine, 7-methylguanosine, 2,2-dimethylguanosine, 5-methylaminoethyluridine, 5-methyloxyuridine, deazanucleotides such as 7-deazaadenosine, 6-azouridine, 6-azocytidine, 6-azothymidine, 5-methyl-2-thiouridine, other thio bases such as 2-thiouridine and 4-thiouridine and 2-thiocytidine, dihydrouridine, pseudouridine, queuosine, archaeosine, naphthyl and substituted naphthyl groups, any O- and N-alkylated purines and pyrimidines such as N6-methyladenosine, 5-methylcarbonylmethyluridine, uridine 5-oxyacetic acid, pyridine-4-one, pyridine-2-one, phenyl and modified phenyl groups such as aminophenol or 2,4,6-trimethoxy benzene, modified cytosines that act as G-clamp nucleotides, 8-substituted adenines and guanines, 5-substituted uracils and thymines, azapyrimidines, carboxyhydroxyalkyl nucleotides, carboxyalkylaminoalkyl nucleotides, and alkylcarbonylalkylated nucleotides. Modified nucleotides also include those nucleotides that are modified with respect to the sugar moiety, as well as nucleotides having sugars or analogs thereof that are not ribosyl. For example, the sugar moieties may be, or be based on, mannoses, arabinoses, glucopyranoses, galactopyranoses, 4'-thioribose, and other sugars, heterocycles, or carbocycles.

The term nucleotide is also meant to include what are known in the art as universal bases. By way of example, universal bases include but are not limited to 3-nitropyrrole, 5-nitroindole, or nebularine. The term "nucleotide" is also meant to include the N3' to P5' phosphoramidate, resulting from the substitution of a ribosyl 3' oxygen with an amine group.

Further, the term nucleotide also includes those species that have a detectable label, such as for example a radioactive or fluorescent moiety, or mass label attached to the nucleotide.

5

Polynucleotide

The term “polynucleotide” refers to polymers of nucleotides, and includes but is not limited to DNA, RNA, DNA/RNA hybrids including polynucleotide chains of regularly and/or irregularly alternating deoxyribosyl moieties and ribosyl moieties (i.e., wherein alternate nucleotide units have an –OH, then an –H, then an –OH, then an –H, and so on at the 2’ position of a sugar moiety), and modifications of these kinds of polynucleotides, wherein the attachment of various entities or moieties to the nucleotide units at any position are included.

Polyribonucleotide

The term “polyribonucleotide” refers to a polynucleotide comprising two or more modified or unmodified ribonucleotides and/or their analogs. The term “polyribonucleotide” is used interchangeably with the term “oligoribonucleotide.”

Ribonucleotide and ribonucleic acid

The term “ribonucleotide” and the phrase “ribonucleic acid” (RNA), refer to a modified or unmodified nucleotide or polynucleotide comprising at least one ribonucleotide unit. A ribonucleotide unit comprises an hydroxyl group attached to the 2’ position of a ribosyl moiety that has a nitrogenous base attached in N-glycosidic linkage at the 1’ position of a ribosyl moiety, and a moiety that either allows for linkage to another nucleotide or precludes linkage.

Detailed Description of the Invention

The present invention is directed to improving the efficiency of gene silencing by siRNA. Through the inclusion of multiple siRNA sequences that are targeted to a particular gene and/or selecting an siRNA sequence based on certain defined criteria, improved efficiency may be achieved.

The present invention will now be described in connection with preferred embodiments. These embodiments are presented in order to aid in an understanding of the present invention and are not intended, and should not be construed, to limit the invention in any way. All alternatives, modifications and equivalents that may
5 become apparent to those of ordinary skill upon reading this disclosure are included within the spirit and scope of the present invention.

Furthermore, this disclosure is not a primer on RNA interference. Basic concepts known to persons skilled in the art have not been set forth in detail.
10

Optimizing siRNA

According to one embodiment, the present invention provides a method for improving the effectiveness of gene silencing for use to silence a particular gene through the selection of an optimal siRNA. An siRNA selected according to this
15 method may be used individually, or in conjunction with the first embodiment, *i.e.*, with one or more other siRNAs, each of which may or may not be selected by this criteria in order to maximize their efficiency.

The degree to which it is possible to select an siRNA for a given mRNA that
20 maximizes these criteria will depend on the sequence of the mRNA itself. However, the selection criteria will be independent of the target sequence. According to this method, an siRNA is selected for a given gene by using a rational design. That said, rational design can be described in a variety of ways. Rational design is, in simplest terms, the application of a proven set of criteria that enhance the probability of
25 identifying a functional or hyperfunctional siRNA. In one method, rationally designed siRNA can be identified by maximizing one or more of the following criteria:

1. A low GC content, preferably between about 30 –52%.
- 30 2. At least 2, preferably at least 3 A or U bases at positions 15- 19 of the siRNA on the sense strand.
3. An A base at position 19 of the sense strand.
4. An A base at position 3 of the sense strand.
5. A U base at position 10 of the sense strand.

6. An A base at position 14 of the sense strand.
7. A base other than C at position 19 of the sense strand.
8. A base other than G at position 13 of the sense strand.
9. A T_m, which refers to the character of the internal repeat that results in inter- or intramolecular structures for one strand of the duplex, that is preferably not stable at greater than 50°C, more preferably not stable at greater than 37°C, even more preferably not stable at greater than 30°C and most preferably not stable at greater than 20°C.
10. A base other than U at position 5 of the sense strand.
11. A base other than A at position 11 of the sense strand.

Criteria 5, 6, 10 and 11 are minor criteria, but are nonetheless desirable. Accordingly, preferably an siRNA will satisfy as many of the aforementioned criteria as possible, more preferably at least 1 – 4 and 7-9, and most preferably all of the criteria

With respect to the criteria, GC content, as well as a high number of AU in positions 15-19, may be important for easement of the unwinding of double stranded siRNA duplex. Duplex unwinding has been shown to be crucial for siRNA functionality *in vivo*.

With respect to criterion 9, the internal structure is measured in terms of the melting temperature of the single strand of siRNA, which is the temperature at which 50% of the molecules will become denatured. With respect to criteria 2 – 8 and 10 – 11, the positions refer to sequence positions on the sense strand, which is the strand that is identical to the mRNA.

In one preferred embodiment, at least criteria 1 and 8 are satisfied. In another preferred embodiment, at least criteria 7 and 8 are satisfied. In still another preferred embodiment, at least criteria 1, 8 and 9 are satisfied.

It should be noted that all of the aforementioned criteria regarding sequence position specifics are with respect to the 5' end of the sense strand. Reference is made to the sense strand, because most databases contain information that describes

the information of the mRNA. Because according to the present invention a chain can be from 18 to 30 bases in length, and the aforementioned criteria assumes a chain 19 base pairs in length, it is important to keep the aforementioned criteria applicable to the correct bases.

5

When there are only 18 bases, the base pair that is not present is the base pair that is located at the 3' of the sense strand. When there are twenty to thirty bases present, then additional bases are added at the 5' end of the sense chain and occupy positions -1 to -11. Accordingly, with respect to SEQ. ID NO. 0001.

10 NNANANNNNUCNAANNNA and SEQ. ID NO. 0028. GUCNNANANNNNUCNAANNNA, both would have A at position 3, A at position 5, U at position 10, C at position 11, A and position 13; A and position 14 and A at position 19. However, SEQ. ID NO. 0028 would also have C at position -1, U at position -2 and G at position -3.

15

For a 19 base pair siRNA, an optimal sequence of one of the strands may be represented below, where N is any base, A, C, G, or U:

SEQ. ID NO. 0001. NNANANNNNUCNAANNNA
 20 SEQ. ID NO. 0002. NNANANNNNUGNAANNNA
 SEQ. ID NO. 0003. NNANANNNNUUNAANNNA
 SEQ. ID NO. 0004. NNANANNNNUCNCANNNA
 SEQ. ID NO. 0005. NNANANNNNUGNCANNNA
 SEQ. ID NO. 0006. NNANANNNNUUNCANNNA
 25 SEQ. ID NO. 0007. NNANANNNNUCNUANNNA
 SEQ. ID NO. 0008. NNANANNNNUGNUANNNA
 SEQ. ID NO. 0009. NNANANNNNUUNUANNNA
 SEQ. ID NO. 0010. NNANCNNNNUCNAANNNA
 SEQ. ID NO. 0011. NNANCNNNNUGNAANNNA
 30 SEQ. ID NO. 0012. NNANCNNNNUUNAANNNA
 SEQ. ID NO. 0013. NNANCNNNNUCNCANNNA
 SEQ. ID NO. 0014. NNANCNNNNUGNCANNNA
 SEQ. ID NO. 0015. NNANCNNNNUUNCANNNA
 SEQ. ID NO. 0016. NNANCNNNNUCNUANNNA

- SEQ. ID NO. 0017. NNANCNNNNUGNUANNNA
 SEQ. ID NO. 0018. NNANCNNNNUUNUANNNA
 SEQ. ID NO. 0019. NNANGNNNNUCNAANNNA
 SEQ. ID NO. 0020. NNANGNNNNUGNAANNNA
 5 SEQ. ID NO. 0021. NNANGNNNNUUNAANNNA
 SEQ. ID NO. 0022. NNANGNNNNUCNCANNNA
 SEQ. ID NO. 0023. NNANGNNNNUGNCANNNA
 SEQ. ID NO. 0024. NNANGNNNNUUNCANNNA
 SEQ. ID NO. 0025. NNANGNNNNUCNUANNNA
 10 SEQ. ID NO. 0026. NNANGNNNNUGNUANNNA
 SEQ. ID NO. 0027. NNANGNNNNUNUANNNA

In one embodiment, the sequence used as an siRNA is selected by choosing the siRNA that score highest according to one of the following seven algorithms that
 15 are represented by Formulas I - VII:

Formula I

Relative functionality of siRNA = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19})$
 $+ (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$
 20

Formula II

Relative functionality of siRNA = $-(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2$
 $+ (A_3)$

25 Formula III

Relative functionality of siRNA = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3$

Formula IV

Relative functionality of siRNA =
 30 $-GC/2 + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$

Formula V

Relative functionality of siRNA = $-(G_{13})^3 - (C_{19}) + (A_{19})^2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$

Formula VI

5 Relative functionality of siRNA = $-(G_{13})^3 - (C_{19}) + (A_{19})^2 + (A_3)$

Formula VII

Relative functionality of siRNA = $-(GC/2) + (AU_{15-19})/2 - (T_{m20^\circ C})^1 - (G_{13})^3 - (C_{19}) + (A_{19})^3 + (A_3)^3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2$

10

In Formulas I – VII:

wherein A_{19} = 1 if A is the base at position 19 on the sense strand, otherwise its value is 0,

AU_{15-19} = 0 – 5 depending on the number of A or U bases on the sense strand

15 at

positions 15 – 19;

G_{13} = 1 if G is the base at position 13 on the sense strand, otherwise its value is 0;

C_{19} = 1 if C is the base at position 19 of the sense strand, otherwise its value is

20 0;

GC = the number of G and C bases in the entire sense strand;

$T_{m\ 20^\circ C}$ = 1 if the T_m is greater than $20^\circ C$;

A_3 = 1 if A is the base at position 3 on the sense strand, otherwise its value is 0;

U_{10} = 1 if U is the base at position 10 on the sense strand, otherwise its value is

25 0;

A_{14} = 1 if A is the base at position 14 on the sense strand, otherwise its value is 0;

U_5 = 1 if U is the base at position 5 on the sense strand, otherwise its value is 0; and

30 A_{11} = 1 if A is the base at position 11 of the sense strand, otherwise its value is 0.

Formulas I – VII provide relative information regarding functionality. When the values for two sequences are compared for a given formula, the relative

functionality is ascertained; a higher positive number indicates a greater functionality. For example, in many applications a value of 5 or greater is beneficial.

5 Additionally, in many applications, more than one of these formulas would
provide useful information as to the relative functionality of potential siRNA
sequences. However, it is beneficial to have more than one type of formula, because
not every formula will be able to help to differentiate among potential siRNA
sequences. For example, in particularly high GC mRNAs, formulas that take that
parameter into account would not be useful and application of formulas that lack GC
10 elements (*e.g.*, formulas V and VI) might provide greater insights into duplex
functionality. Similarly, formula II might be used in situations where hairpin
structures are not observed in duplexes, and formula IV might be applicable for
sequences that have higher AU content. Thus, one may consider a particular sequence
in light of more than one or even all of these algorithms to obtain the best
15 differentiation among sequences. In some instances, application of a given algorithm
may identify an unusually large number of potential siRNA sequences, and in those
cases, it may be appropriate to re-analyze that sequence with a second algorithm that
is, for instance, more stringent. Alternatively, it is conceivable that analysis of a
sequence with a given formula yields no acceptable siRNA sequences (*i.e.* low
20 SMARTscores™). In this instance, it may be appropriate to re-analyze that sequences
with a second algorithm that is, for instance, less stringent. In still other instances,
analysis of a single sequence with two separate formulas may give rise to conflicting
results (*i.e.* one formula generates a set of siRNA with high SMARTscores™ while
the other formula identifies a set of siRNA with low SMARTscores™). In these
25 instances, it may be necessary to determine which weighted factor(s) (*e.g.* GC
content) are contributing to the discrepancy and assessing the sequence to decide
whether these factors should or should not be included. Alternatively, the sequence
could be analyzed by a third, fourth, or fifth algorithm to identify a set of rationally
designed siRNA.

30

The above-referenced criteria are particularly advantageous when used in
combination with pooling techniques as depicted in Table I:

Table I

Criteria	Functional Probability					
	Oligos			Pools		
	>95%	>80%	<70%	>95%	>80%	<70%
Current	33.0	50.0	23.0	79.5	97.3	0.3
New	50.0	88.5	8.0	93.8	99.98	0.005
(GC)	28.0	58.9	36.0	72.8	97.1	1.6

The term “current” refers to Tuschl’s conventional siRNA parameters (Elbashir, S.M. et al. (2002) “Analysis of gene function in somatic mammalian cells using small interfering RNAs” Methods 26: 199-213). “New” refers to the design parameters described in Formulas I-VII. “GC” refers to criteria that select siRNA solely on the basis of GC content.

As Table I indicates, when more functional siRNA duplexes are chosen, siRNAs that produce <70% silencing drops from 23% to 8% and the number of siRNA duplexes that produce >80% silencing rises from 50% to 88.5%. Further, of the siRNA duplexes with >80% silencing, a larger portion of these siRNAs actually silence >95% of the target expression (the new criteria increases the portion from 33% to 50%). Using this new criteria in pooled siRNAs, shows that, with pooling, the amount of silencing >95% increases from 79.5% to 93.8% and essentially eliminates any siRNA pool from silencing less than 70%.

Table II similarly shows the particularly beneficial results of pooling in combination with the aforementioned criteria. However, Table II, which takes into account each of the aforementioned variables, demonstrates even a greater degree of improvement in functionality.

Table II

	Functional Probability					
	Oligos			Pools		
	Functional	Average	Non-functional	Functional	Average	Non-functional
Random	20	40	50	67	97	3

Criteria 1	52	99	0.1	97	93	0.0040
Criteria 4	89	99	0.1	99	99	0.0000

The terms "functional," "Average," and "Non-functional" refer to siRNA that exhibit >80%, >50%, and <50% functionality, respectively. Criteria 1 and 4 refer to specific criteria described above.

5

The above-described algorithms may be used with or without a computer program that allows for the inputting of the sequence of the mRNA and automatically outputs the optimal siRNA. The computer program may, for example, be accessible from a local terminal or personal computer, over an internal network or over the Internet.

10

In addition to the formulas above, more detailed algorithms may be used for selecting siRNA. Preferably, at least one RNA duplex of between 18 and 30 base pairs is selected such that it is optimized according to a formula selected from:

15

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

20

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$

25

wherein

30

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;

$A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;

$A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
5 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is
10 present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
15 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
20 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is
present or the sense strand is only 18 base pairs in length, its value is 0;

$G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
25 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is
present or the sense strand is only 18 base pairs in length, its value is 0;

30 $U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;

$U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
5 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

10 GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand, or within positions 15 – 18 if the sense strand is only 18 base pairs in length;

GC_{total} = the number of G and C bases in the sense strand;

$T_m = 100$ if the siRNA oligo has the internal repeat longer than 4 base pairs, otherwise its value is 0; and

15 X = the number of times that the same nucleotide repeats four or more times in a row.

20 The above formulas VIII and IX, as well as formulas I – VII, provide methods for selecting siRNA in order to increase the efficiency of gene silencing. A subset of variables of any of the formulas may be used, though when fewer variables are used, the optimization hierarchy becomes less reliable.

25 With respect to the variables of the above-referenced formulas, a single letter of A or C or G or U followed by a subscript refers to a binary condition. The binary condition is that either the particular base is present at that particular position (wherein the value is “1”) or the base is not present (wherein the value is “0”). Because position 19 is optional, *i.e.* there might be only 18 base pairs, when there are only 18 base pairs, any base with a subscript of 19 in the formulas above would have a zero value for that parameter. Before or after each variable is a number followed by
 30 *, which indicates that the value of the variable is to be multiplied or weighed by that number.

The numbers preceding the variables A, or G, or C, or U in Formulas VIII and IX (or after the variables in Formula I - VII) were determined by comparing the

difference in the frequency of individual bases at different positions in functional siRNA and total siRNA. Specifically, the frequency in which a given base was observed at a particular position in functional groups was compared with the frequency that that same base was observed in the total, randomly selected siRNA set.

5 If the absolute value of the difference between the functional and total values was found to be greater than 6%, that parameter was included in the equation. Thus for instance, if the frequency of finding a "G" at position 13 (G_{13}) is found to be 6% in a given functional group, and the frequency of G_{13} in the total population of siRNAs is 20%, the difference between the two values is $6\% - 20\% = -14\%$. As the absolute value

10 is greater than six (6), this factor (-14) is included in the equation. Thus in Formula VIII, in cases where the siRNA under study has a G in position 13, the accrued value is $(-14) * (1) = -14$. In contrast, when a base other than G is found at position 13, the accrued value is $(-14) * (0) = 0$.

15 When developing a means to optimize siRNAs, the inventors observed that a bias toward low internal thermodynamic stability of the duplex at the 5'-antisense (AS) end is characteristic of naturally occurring miRNA precursors. The inventors extended this observation to siRNAs for which functionality had been assessed in tissue culture.

20 With respect to the parameter GC_{15-19} , a value of 0 – 5 will be ascribed depending on the number of G or C bases at positions 15 to 19. If there are only 18 base pairs, the value is between 0 and 4.

25 With respect to the criterion GC_{total} content, a number from 0 – 30 will be ascribed, which correlates to the total number of G and C nucleotides on the sense strand, excluding overhangs. Without wishing to be bound by any one theory, it is postulated that the significance of the GC content (as well as AU content at positions 15-19, which is a parameter for formulas III – VII) relates to the easement of the

30 unwinding of a double-stranded siRNA duplex. Duplex unwinding is believed to be crucial for siRNA functionality *in vivo* and overall low internal stability, especially low internal stability of the first unwound base pair is believed to be important to maintain sufficient processivity of RISC complex-induced duplex unwinding. If the duplex has 19 base pairs, those at positions 15-19 on the sense strand will unwind first

if the molecule exhibits a sufficiently low internal stability at that position. As persons skilled in the art are aware, RISC is a complex of approximately twelve proteins; Dicer is one, but not the only, helicase within this complex. Accordingly, although the GC parameters are believed to relate to activity with Dicer, they are also
5 important for activity with other RISC proteins.

The value of the parameter T_m is 0 when there are no internal repeats longer than (or equal to) four base pairs present in the siRNA duplex; otherwise the value is 1. Thus for example, if the sequence ACGUACGU, or any other four nucleotide (or
10 more) palindrome exists within the structure, the value will be one (1). Alternatively if the structure ACGGACG, or any other 3 nucleotide (or less) palindrome exists, the value will be zero (0).

The variable "X" refers to the number of times that the same nucleotide occurs
15 contiguously in a stretch of four or more units. If there are, for example, four contiguous As in one part of the sequence and elsewhere in the sequence four contiguous Cs, $X=2$. Further, if there are two separate contiguous stretches of four of the same nucleotides or eight or more of the same nucleotides in a row, then $X=2$. However, X does not increase for five, six or seven contiguous nucleotides.

20
Again, when applying Formula VIII or Formula IX to a given mRNA, (the "target RNA" or "target molecule"), one may use a computer program to evaluate the criteria for every sequence of 18 – 30 base pairs or only sequences of a fixed length, e.g., 19 base pairs. Preferably the computer program is designed such that it provides
25 a report ranking of all of the potential siRNAs between 18 and 30 base pairs, ranked according to which sequences generate the highest value. A higher value refers to a more efficient siRNA for a particular target gene. The computer program that may be used, may be developed in any computer language that is known to be useful for scoring nucleotide sequences, or it may be developed with the assistance of
30 commercially available product such as Microsoft's product .net. Additionally, rather than run every sequence through one and/or another formula, one may compare a subset of the sequences, which may be desirable if for example only a subset are available. For instance, it may be desirable to first perform a BLAST (Basic Local Alignment Search Tool) search and to identify sequences that have no homology to

other targets. Alternatively, it may be desirable to scan the sequence and to identify regions of moderate GC context, then perform relevant calculations using one of the above-described formulas on these regions. These calculations can be done manually or with the aid of a computer.

5

As with Formulas I – VII, either Formula VIII or Formula IX may be used for a given mRNA target sequence. However, it is possible that according to one or the other formula more than one siRNA will have the same value. Accordingly, it is beneficial to have a second formula by which to differentiate sequences. Formula IX was derived in a similar fashion as Formula VIII, yet used a larger data set and thus yields sequences with higher statistical correlations to highly functional duplexes. The sequence that has the highest value ascribed to it may be referred to as a “first optimized duplex.” The sequence that has the second highest value ascribed to it may be referred to as a “second optimized duplex.” Similarly, the sequences that have the third and fourth highest values ascribed to them may be referred to as a third optimized duplex and a fourth optimized duplex, respectively. When more than one sequence has the same value, each of them may, for example, be referred to as first optimized duplex sequences or co-first optimized duplexes.

20 SiRNA sequences identified using Formula VIII are contained within the enclosed compact disks. The data included on the enclosed compact disks is described more fully below. The sequences identified by Formula VIII that are disclosed in the compact disks may be used in gene silencing applications.

25 It should be noted that for Formulas VIII and IX all of the aforementioned criteria are identified as positions on the sense strand when oriented in the 5' to 3' direction as they are identified in connection with Formulas I – VII unless otherwise specified.

30 Formulas I - IX, may be used to select or to evaluate one, or more than one, siRNA in order to optimize silencing. Preferably, at least two optimized siRNAs that have been selected according to at least one of these formulas are used to silence a gene, more preferably at least three and most preferably at least four. The siRNAs may be used individually or together in a pool or kit. Further, they may be applied to

a cell simultaneously or separately. Preferably, the at least two siRNAs are applied simultaneously. Pools are particularly beneficial for many research applications. However, for therapeutics, it may be more desirable to employ a single hyperfunctional siRNA as described elsewhere in this application.

5

When planning to conduct gene silencing, and it is necessary to choose between two or more siRNAs, one should do so by comparing the relative values when the siRNA are subjected to one of the formulas above. In general a higher scored siRNA should be used.

10

Useful applications include, but are not limited to, target validation, gene functional analysis, research and drug discovery, gene therapy and therapeutics. Methods for using siRNA in these applications are well known to persons of skill in the art.

15

Because the ability of siRNA to function is dependent on the sequence of the RNA and not the species into which it is introduced, the present invention is applicable across a broad range of species, including but not limited to all mammalian species, such as humans, dogs, horses, cats, cows, mice, hamsters, chimpanzees and gorillas, as well as other species and organisms such as bacteria, viruses, insects, plants and *C. elegans*.

20

The present invention is also applicable for use for silencing a broad range of genes, including but not limited to the roughly 45,000 genes of a human genome, and has particular relevance in cases where those genes are associated with diseases such as diabetes, Alzheimer's, cancer, as well as all genes in the genomes of the aforementioned organisms.

25

The siRNA selected according to the aforementioned criteria or one of the aforementioned algorithms are also, for example, useful in the simultaneous screening and functional analysis of multiple genes and gene families using high throughput strategies, as well as in direct gene suppression or silencing.

30

Development of the Algorithms

To identify siRNA sequence features that promote functionality and to quantify the importance of certain currently accepted conventional factors—such as G/C content and target site accessibility—the inventors synthesized an siRNA panel consisting of 270 siRNAs targeting three genes, Human Cyclophilin, Firefly

5 Luciferase, and Human DBI. In all three cases, siRNAs were directed against specific regions of each gene. For Human Cyclophilin and Firefly Luciferase, ninety siRNAs were directed against a 199 bp segment of each respective mRNA. For DBI, 90 siRNAs were directed against a smaller, 109 base pair region of the mRNA. The sequences to which the siRNAs were directed are provided below.

10

It should be noted that in certain sequences, “t” is present. This is because many databases contain information in this manner. However, the t denotes a uracil residue in mRNA and siRNA. Any algorithm will, unless otherwise specified, process a t in a sequence as a u.

15

Human cyclophilin: 193—390, M60857

SEQ. ID NO. 29:

gttccaaaacagtggataattttgtggccttagctacaggagagaaaggatttggctacaaaacagcaaattccatcgtgt
aatcaaggacttcatgatccagggcgagacttcaccaggggagatggcacaggaggaaagagcatctacggtgagcg
20 cttccccgatgagaacttcaaactgaagcactacgggcctggctggg

Firefly luciferase: 1434—1631, U47298 (pGL3, Promega)

SEQ. ID NO. 30:

tgaactcccgccgctgtgttttggagcacggaaagacgatgacggaaaaagagatcgtggattacgtcgccagtca
25 agtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagtaccgaaaggcttaccgaaaactcgacg
caagaaaaatcagagagatcctcataaaggccaagaagg

DBI, NM_020548 (202-310) (every position)

SEQ. ID NO. 0031:

30 acgggcaaggccaagtgggatgcctggaatgagctgaaagggacttccaaggaagatgcatgaaagcttacatcaaca
aagtagaagagctaaagaaaaaacg

A list of the siRNAs appears in Table III (see Examples Section, Example II)

The set of duplexes was analyzed to identify correlations between siRNA functionality and other biophysical or thermodynamic properties. When the siRNA panel was analyzed in functional and non-functional subgroups, certain nucleotides were much more abundant at certain positions in functional or non-functional groups.

5 More specifically, the frequency of each nucleotide at each position in highly functional siRNA duplexes was compared with that of nonfunctional duplexes in order to assess the preference for or against any given nucleotide at every position. These analyses were used to determine important criteria to be included in the siRNA algorithms (Formulas VIII and IX).

10

The data set was also analyzed for distinguishing biophysical properties of siRNAs in the functional group, such as optimal percent of GC content, propensity for internal structures and regional thermodynamic stability. Of the presented criteria, several are involved in duplex recognition, RISC activation/duplex unwinding, and target cleavage catalysis.

15

The original data set that was the source of the statistically derived criteria is shown in **Figure 2**. Additionally, this figure shows that random selection yields siRNA duplexes with unpredictable and widely varying silencing potencies as measured in tissue culture using HEK293 cells. In the figure, duplexes are plotted such that each x-axis tick-mark represents an individual siRNA, with each subsequent siRNA differing in target position by two nucleotides for Human Cyclophilin and Firefly Luciferase, and by one nucleotide for Human DBI. Furthermore, the y-axis denotes the level of target expression remaining after transfection of the duplex into cells and subsequent silencing of the target.

20

SiRNA identified and optimized in this document work equally well in a wide range of cell types. **Figure 3a** shows the evaluation of thirty siRNAs targeting the DBI gene in three cell lines derived from different tissues. Each DBI siRNA displays very similar functionality in HEK293 (ATCC, CRL-1573, human embryonic kidney), HeLa (ATCC, CCL-2, cervical epithelial adenocarcinoma) and DU145 (HTB-81, prostate) cells as determined by the B-DNA assay. Thus, siRNA functionality is determined by the primary sequence of the siRNA and not by the intracellular environment. Additionally, it should be noted that although the present invention

30

provides for a determination of the functionality of siRNA for a given target, the same siRNA may silence more than one gene. For example, the complementary sequence of the silencing siRNA may be present in more than one gene. Accordingly, in these circumstances, it may be desirable not to use the siRNA with highest SMARTscore™.

5 In such circumstances, it may be desirable to use the siRNA with the next highest SMARTscore™.

To determine the relevance of G/C content in siRNA function, the G/C content of each duplex in the panel was calculated and the functional classes of siRNAs

10 (<F50, \geq F50, \geq F80, \geq F95 where F refers to the percent gene silencing) were sorted accordingly. The majority of the highly-functional siRNAs (\geq F95) fell within the G/C content range of 36%—52% (Figure 3B). Twice as many non-functional (< F50) duplexes fell within the high G/C content groups (>57% GC content) compared to the 36%—52% group. The group with extremely low GC content (26% or less)

15 contained a higher proportion of non-functional siRNAs and no highly-functional siRNAs. The G/C content range of 30%—52% was therefore selected as Criterion I for siRNA functionality, consistent with the observation that a G/C range 30%—70% promotes efficient RNAi targeting. Application of this criterion alone provided only a marginal increase in the probability of selecting functional siRNAs from the panel:

20 selection of F50 and F95 siRNAs was improved by 3.6% and 2.2%, respectively. The siRNA panel presented here permitted a more systematic analysis and quantification of the importance of this criterion than that used previously.

A relative measure of local internal stability is the A/U base pair (bp) content; therefore, the frequency of A/U bp was determined for each of the five terminal

25 positions of the duplex (5' sense (S)/5' antisense (AS)) of all siRNAs in the panel. Duplexes were then categorized by the number of A/U bp in positions 1—5 and 15—19 of the sense strand. The thermodynamic flexibility of the duplex 5'-end (positions 1—5; S) did not appear to correlate appreciably with silencing potency, while that of

30 the 3'-end (positions 15—19; S) correlated with efficient silencing. No duplexes lacking A/U bp in positions 15—19 were functional. The presence of one A/U bp in this region conferred some degree of functionality, but the presence of three or more A/Us was preferable and therefore defined as Criterion II. When applied to the test panel, only a marginal increase in the probability of functional siRNA selection was

achieved: a 1.8% and 2.3% increase for F50 and F95 duplexes, respectively (Table IV).

The complementary strands of siRNAs that contain internal repeats or
5 palindromes may form internal fold-back structures. These hairpin-like structures exist in equilibrium with the duplexed form effectively reducing the concentration of functional duplexes. The propensity to form internal hairpins and their relative stability can be estimated by predicted melting temperatures. High T_m reflects a tendency to form hairpin structures. Lower T_m values indicate a lesser tendency to
10 form hairpins. When the functional classes of siRNAs were sorted by T_m (Figure 3c), the following trends were identified: duplexes lacking stable internal repeats were the most potent silencers (no F95 duplex with predicted hairpin structure $T_m > 60^\circ\text{C}$). In contrast, about 60% of the duplexes in the groups having internal hairpins with calculated T_m values less than 20°C were F80. Thus, the stability of internal repeats
15 is inversely proportional to the silencing effect and defines Criterion III (predicted hairpin structure $T_m \leq 20^\circ\text{C}$).

Sequence-based determinants of siRNA functionality

When the siRNA panel was sorted into functional and non-functional groups,
20 the frequency of a specific nucleotide at each position in a functional siRNA duplex was compared with that of a nonfunctional duplex in order to assess the preference for or against a certain nucleotide. Figure 4 shows the results of these queries and the subsequent resorting of the data set (from Figure 2). The data is separated into two sets: those duplexes that meet the criteria, a specific nucleotide in a certain position -
25 grouped on the left (Selected) and those that do not - grouped on the right (Eliminated). The duplexes are further sorted from most functional to least functional with the y-axis of Figure 4a-e representing the % expression *i.e.* the amount of silencing that is elicited by the duplex (Note: each position on the X-axis represents a different duplex). Statistical analysis revealed correlations between silencing and
30 several sequence-related properties of siRNAs. Figure 4 and Table IV show quantitative analysis for the following five sequence-related properties of siRNA: (A) an A at position 19 of the sense strand; (B) an A at position 3 of the sense strand; (C) a U at position 10 of the sense strand; (D) a base other than G at position 13 of the sense strand; and (E) a base other than C at position 19 of the sense strand.

When the siRNAs in the panel were evaluated for the presence of an A at position 19 of the sense strand, the percentage of non-functional duplexes decreased from 20% to 11.8%, and the percentage of F95 duplexes increased from 21.7% to 29.4% (**Table IV**). Thus, the presence of an A in this position defined Criterion IV.

Another sequence-related property correlated with silencing was the presence of an A in position 3 of the sense strand (**Figure 4b**). Of the siRNAs with A3, 34.4% were F95, compared with 21.7% randomly selected siRNAs. The presence of a U base in position 10 of the sense strand exhibited an even greater impact (**Figure 4c**). Of the duplexes in this group, 41.7% were F95. These properties became criteria V and VI, respectively.

Two negative sequence-related criteria that were identified also appear on **Figure 4**. The absence of a G at position 13 of the sense strand, conferred a marginal increase in selecting functional duplexes (**Figure 4d**). Similarly, lack of a C at position 19 of the sense strand also correlated with functionality (**Figure 4e**). Thus, among functional duplexes, position 19 was most likely occupied by A, and rarely occupied by C. These rules were defined as criteria VII and VIII, respectively.

Application of each criterion individually provided marginal but statistically significant increases in the probability of selecting a potent siRNA. Although the results were informative, the inventors sought to maximize potency and therefore consider multiple criteria or parameters. Optimization is particularly important when developing therapeutics. Interestingly, the probability of selecting a functional siRNA based on each thermodynamic criteria was 2%—4% higher than random, but 4%—8% higher for the sequence-related determinates. Presumably, these sequence-related increases reflect the complexity of the RNAi mechanism and the multitude of protein-RNA interactions that are involved in RNAi-mediated silencing.

Table IV

Criterion	% Functional		Improvement over Random
I. 30%—52% G/C content	< F50	16.4%	-3.6%
	≥ F50	83.6%	3.6%
	≥ F80	60.4%	4.3%
	≥ F95	23.9%	2.2%
II. At least 3 A/U bases at positions 15—19 of the sense strand	< F50	18.2%	-1.8%
	≥ F50	81.8%	1.8%
	≥ F80	59.7%	3.6%
	≥ F95	24.0%	2.3%
III. Absence of internal repeats, as measured by T_m of secondary structure $\leq 20^\circ\text{C}$	< F50	16.7%	-3.3%
	≥ F50	83.3%	3.3%
	≥ F80	61.1%	5.0%
	≥ F95	24.6%	2.9%
IV. An A base at position 19 of the sense strand	< F50	11.8%	-8.2%
	≥ F50	88.2%	8.2%
	≥ F80	75.0%	18.9%
	≥ F95	29.4%	7.7%
V. An A base at position 3 of the sense strand	< F50	17.2%	-2.8%
	≥ F50	82.8%	2.8%
	≥ F80	62.5%	6.4%
	≥ F95	34.4%	12.7%
VI. A U base at position 10 of the sense strand	< F50	13.9%	-6.1%
	≥ F50	86.1%	6.1%
	≥ F80	69.4%	13.3%
	≥ F95	41.7%	20%
VII. A base other than C at position 19 of the sense strand	< F50	18.8%	-1.2%
	≥ F50	81.2%	1.2%
	≥ F80	59.7%	3.6%
	≥ F95	24.2%	2.5%
VIII. A base other than G at position 13 of the sense strand	< F50	15.2%	-4.8%
	≥ F50	84.8%	4.8%
	≥ F80	61.4%	5.3%
	≥ F95	26.5%	4.8%

The siRNA selection algorithm

In an effort to improve selection further, all identified criteria, including but not limited to those listed in Table IV were combined into the algorithms embodied in Formula VIII and Formula IX. Each siRNA was then assigned a score (referred to as a SMARTscore™) according to the values derived from the formulas. Duplexes that

scored higher than 0 or 20, for Formulas VIII and IX, respectively, effectively selected a set of functional siRNAs and excluded all non-functional siRNAs.

Conversely, all duplexes scoring lower than 0 and 20 according to formulas VIII and

IX, respectively, contained some functional siRNAs but included all non-functional siRNAs. A graphical representation of this selection is shown in **Figure 5**.

5 The methods for obtaining the seven criteria embodied in Table IV are illustrative of the results of the process used to develop the information for Formulas VIII and IX. Thus similar techniques were used to establish the other variables and their multipliers. As described above, basic statistical methods were use to determine the relative values for these multipliers.

10 To determine the value for “Improvement over Random” the difference in the frequency of a given attribute (*e.g.* GC content, base preference) at a particular position is determined between individual functional groups (*e.g.* <F50) and the total siRNA population studied (*e.g.* 270 siRNA molecules selected randomly). Thus, for instance, in Criterion I (30%-52% GC content) members of the <F50 group were
15 observed to have GC contents between 30-52% in 16.4% of the cases. In contrast, the total group of 270 siRNAs had GC contents in this range, 20% of the time. Thus for this particular attribute, there is a small negative correlation between 30%-52% GC content and this functional group (*i.e.* $16.4\% - 20\% = -3.6\%$). Similarly, for Criterion VI, (a “U” at position 10 of the sense strand), the >F95 group contained a “U” at this
20 position 41.7% of the time. In contrast, the total group of 270 siRNAs had a “U” at this position 21.7% of the time, thus the improvement over random is calculated to be 20% (or $41.7\% - 21.7\%$).

Identifying The Average Internal Stability Profile of Strong siRNA

25 In order to identify an internal stability profile that is characteristic of strong siRNA, 270 different siRNAs derived from the cyclophilin B, the diazepam binding inhibitor (DBI), and the luciferase gene were individually transfected into HEK293 cells and tested for their ability to induce RNAi of the respective gene. Based on their performance in the *in vivo* assay, the sequences were then subdivided into three
30 groups, (i) >95% silencing; (ii) 80-95% silencing; and (iii) less than 50% silencing. Sequences exhibiting 51-84% silencing were eliminated from further consideration to reduce the difficulties in identifying relevant thermodynamic patterns.

Following the division of siRNA into three groups, a statistical analysis was performed on each member of each group to determine the average internal stability profile (AISP) of the siRNA. To accomplish this the Oligo 5.0 Primer Analysis Software and other related statistical packages (*e.g.* Excel) were exploited to
5 determine the internal stability of pentamers using the nearest neighbor method described by Freier *et al.*, (1986) *Improved free-energy parameters for predictions of RNA duplex stability*, Proc Natl. Acad. Sci. U. S. A. 83(24): 9373-7. Values for each group at each position were then averaged, and the resulting data were graphed on a linear coordinate system with the Y-axis expressing the ΔG (free energy) values in
10 kcal/mole and the X-axis identifying the position of the base relative to the 5' end.

The results of the analysis identified multiple key regions in siRNA molecules that were critical for successful gene silencing. At the 3'-most end of the sense strand (5'antisense), highly functional siRNA (>95% gene silencing, see **Figure 6a**, >F95)
15 have a low internal stability (AISP of position 19 = ~ -7.6 kcal/mol). In contrast low-efficiency siRNA (*i.e.* those exhibiting less than 50% silencing, <F50) display a distinctly different profile, having high ΔG values (~ -8.4 kcal/mol) for the same position. Moving in a 5' (sense strand) direction, the internal stability of highly efficient siRNA rises (position 12 = ~ -8.3 kcal/mole) and then drops again (position 7
20 = ~ -7.7 kcal/mol) before leveling off at a value of approximately -8.1 kcal/mol for the 5' terminus. SiRNA with poor silencing capabilities show a distinctly different profile. While the AISP value at position 12 is nearly identical with that of strong siRNAs, the values at positions 7 and 8 rise considerably, peaking at a high of ~ -9.0 kcal/mol. In addition, at the 5' end of the molecule the AISP profile of strong and
25 weak siRNA differ dramatically. Unlike the relatively strong values exhibited by siRNA in the >95% silencing group, siRNAs that exhibit poor silencing activity have weak AISP values (-7.6 , -7.5 , and -7.5 kcal/mol for positions 1, 2 and 3 respectively).

Overall the profiles of both strong and weak siRNAs form distinct sinusoidal
30 shapes that are roughly 180° out-of-phase with each other. While these thermodynamic descriptions define the archetypal profile of a strong siRNA, it will likely be the case that neither the ΔG values given for key positions in the profile or the absolute position of the profile along the Y-axis (*i.e.* the ΔG -axis) are absolutes.

Profiles that are shifted upward or downward (*i.e.* having on an average, higher or lower values at every position) but retain the relative shape and position of the profile along the X-axis can be foreseen as being equally effective as the model profile described here. Moreover, it is likely that siRNA that have strong or even stronger gene-specific silencing effects might have exaggerated ΔG values (either higher or lower) at key positions. Thus, for instance, it is possible that the 5'-most position of the sense strand (position 19) could have ΔG values of 7.4 kcal/mol or lower and still be a strong siRNA if, for instance, a G-C \rightarrow G-T/U mismatch were substituted at position 19 and altered duplex stability. Similarly, position 12 and position 7 could have values above 8.3 kcal/mol and below 7.7 kcal/mole, respectively, without abating the silencing effectiveness of the molecule. Thus, for instance, at position 12, a stabilizing chemical modification (*e.g.* a chemical modification of the 2' position of the sugar backbone) could be added that increases the average internal stability at that position. Similarly, at position 7, mismatches similar to those described previously could be introduced that would lower the ΔG values at that position.

Lastly, it is important to note that while functional and non-functional siRNA were originally defined as those molecules having specific silencing properties, both broader or more limiting parameters can be used to define these molecules. As used herein, unless otherwise specified, "non-functional siRNA" are defined as those siRNA that induce less than 50% (<50%) target silencing, "semi-functional siRNA" induce 50-79% target silencing, "functional siRNA" are molecules that induce 80-95% gene silencing, and "highly-functional siRNA" are molecules that induce greater than 95% gene silencing. These definitions are not intended to be rigid and can vary depending upon the design and needs of the application. For instance, it is possible that a researcher attempting to map a gene to a chromosome using a functional assay, may identify an siRNA that reduces gene activity by only 30%. While this level of gene silencing may be "non-functional" for *e.g.* therapeutic needs, it is sufficient for gene mapping purposes and is, under these uses and conditions, "functional." For these reasons, functional siRNA can be defined as those molecules having greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% silencing capabilities at 100nM transfection conditions. Similarly, depending upon the needs of the study and/or application, non-functional and semi-functional siRNA can be defined as

having different parameters. For instance, semi-functional siRNA can be defined as being those molecules that induce 20%, 30%, 40%, 50%, 60%, or 70% silencing at 100nM transfection conditions. Similarly, non-functional siRNA can be defined as being those molecules that silence gene expression by less than 70%, 60%, 50%, 40%, 30%, or less. Nonetheless, unless otherwise stated, the descriptions stated in the “Definitions” section of this text should be applied.

Functional attributes can be assigned to each of the key positions in the AISP of strong siRNA. The low 5' (sense strand) AISP values of strong siRNAs may be necessary for determining which end of the molecule enters the RISC complex. In contrast, the high and low AISP values observed in the central regions of the molecule may be critical for siRNA-target mRNA interactions and product release, respectively.

If the AISP values described above accurately define the thermodynamic parameters of strong siRNA, it would be expected that similar patterns would be observed in strong siRNA isolated from nature. Natural siRNAs exist in a harsh, RNase-rich environment and it can be hypothesized that only those siRNA that exhibit heightened affinity for RISC (*i.e.* siRNA that exhibit an average internal stability profile similar to those observed in strong siRNA) would survive in an intracellular environment. This hypothesis was tested using GFP-specific siRNA isolated from *N. benthamiana*. Llave *et al.* (2002) *Endogenous and Silencing-Associated Small RNAs in Plants*, The Plant Cell 14, 1605-1619, introduced long double-stranded GFP-encoding RNA into plants and subsequently re-isolated GFP-specific siRNA from the tissues. The AISP of fifty-nine of these GFP-siRNA were determined, averaged, and subsequently plotted alongside the AISP profile obtained from the cyclophilin B/DBI/ luciferase siRNA having >90% silencing properties (**Figure 6b**). Comparison of the two groups show that profiles are nearly identical. This finding validates the information provided by the internal stability profiles and demonstrates that: (1) the profile identified by analysis of the cyclophilin B/DBI/ luciferase siRNAs are not gene specific; and (2) AISP values can be used to search for strong siRNAs in a variety of species.

Both chemical modifications and base-pair mismatches can be incorporated into siRNA to alter the duplex's AISP and functionality. For instance, introduction of mismatches at positions 1 or 2 of the sense strand destabilized the 5' end of the sense strand and increases the functionality of the molecule (see Luc, **Figure 7**). Similarly, addition of 2'-O-methyl groups to positions 1 and 2 of the sense strand can also alter the AISP and (as a result) increase both the functionality of the molecule and eliminate off-target effects that results from sense strand homology with the unrelated targets (**Figures 8a, 8b**).

10 Rationale for Criteria in a Biological Context

The fate of siRNA in the RNAi pathway may be described in 5 major steps: (1) duplex recognition and pre-RISC complex formation; (2) ATP-dependent duplex unwinding/strand selection and RISC activation; (3) mRNA target identification; (4) mRNA cleavage, and (5) product release (**Figure 1**). Given the level of nucleic acid-protein interactions at each step, siRNA functionality is likely influenced by specific biophysical and molecular properties that promote efficient interactions within the context of the multi-component complexes. Indeed, the systematic analysis of the siRNA test set identified multiple factors that correlate well with functionality. When combined into a single algorithm, they proved to be very effective in selecting active siRNAs.

The factors described here may also be predictive of key functional associations important for each step in RNAi. For example, the potential formation of internal hairpin structures correlated negatively with siRNA functionality. Complementary strands with stable internal repeats are more likely to exist as stable hairpins thus decreasing the effective concentration of the functional duplex form. This suggests that the duplex is the preferred conformation for initial pre-RISC association. Indeed, although single complementary strands can induce gene silencing, the effective concentration required is at least two orders of magnitude higher than that of the duplex form.

siRNA-pre-RISC complex formation is followed by an ATP-dependent duplex unwinding step and "activation" of the RISC. The siRNA functionality was shown to correlate with overall low internal stability of the duplex and low internal stability of

the 3' sense end (or differential internal stability of the 3' sense compare to the 5' sense strand), which may reflect strand selection and entry into the RISC. Overall duplex stability and low internal stability at the 3' end of the sense strand were also correlated with siRNA functionality. Interestingly, siRNAs with very high and very low overall stability profiles correlate strongly with non-functional duplexes. One interpretation is that high internal stability prevents efficient unwinding while very low stability reduces siRNA target affinity and subsequent mRNA cleavage by the RISC.

Several criteria describe base preferences at specific positions of the sense strand and are even more intriguing when considering their potential mechanistic roles in target recognition and mRNA cleavage. Base preferences for A at position 19 of the sense strand but not C, are particularly interesting because they reflect the same base preferences observed for naturally occurring miRNA precursors. That is, among the reported miRNA precursor sequences 75% contain a U at position 1 which corresponds to an A in position 19 of the sense strand of siRNAs, while G was under-represented in this same position for miRNA precursors. These observations support the hypothesis that both miRNA precursors and siRNA duplexes are processed by very similar if not identical protein machinery. The functional interpretation of the predominance of a U/A base pair is that it promotes flexibility at the 5' antisense ends of both siRNA duplexes and miRNA precursors and facilitates efficient unwinding and selective strand entrance into an activated RISC.

Among the criteria associated with base preferences that are likely to influence mRNA cleavage or possibly product release, the preference for U at position 10 of the sense strand exhibited the greatest impact, enhancing the probability of selecting an F80 sequence by 13.3%. Activated RISC preferentially cleaves target mRNA between nucleotides 10 and 11 relative to the 5' end of the complementary targeting strand. Therefore, it may be that U, the preferred base for most endoribonucleases, at this position supports more efficient cleavage. Alternatively, a U/A bp between the targeting siRNA strand and its cognate target mRNA may create an optimal conformation for the RISC-associated "slicing" activity.

Pooling

According to another embodiment, the present invention provides a pool of at least two siRNAs, preferably in the form of a kit or therapeutic reagent, wherein one strand of each of the siRNAs, the sense strand comprises a sequence that is
5 substantially similar to a sequence within a target mRNA. The opposite strand, the antisense strand, will preferably comprise a sequence that is substantially complementary to that of the target mRNA. More preferably, one strand of each siRNA will comprise a sequence that is identical to a sequence that is contained in the target mRNA. Most preferably, each siRNA will be 19 base pairs in length, and one
10 strand of each of the siRNAs will be 100% complementary to a portion of the target mRNA.

By increasing the number of siRNAs directed to a particular target using a pool or kit, one is able both to increase the likelihood that at least one siRNA with
15 satisfactory functionality will be included, as well as to benefit from additive or synergistic effects. Further, when two or more siRNAs directed against a single gene do not have satisfactory levels of functionality alone, if combined, they may satisfactorily promote degradation of the target messenger RNA and successfully inhibit translation. By including multiple siRNAs in the system, not only is the
20 probability of silencing increased, but the economics of operation are also improved when compared to adding different siRNAs sequentially. This effect is contrary to the conventional wisdom that the concurrent use of multiple siRNA will negatively impact gene silencing (*e.g.* Holen, T. *et al.* (2003) "Similar behavior of single strand and double strand siRNAs suggests they act through a common RNAi pathway."
25 NAR 31: 2401-21407).

In fact, when two siRNAs were pooled together, 54% of the pools of two siRNAs induced more than 95% gene silencing. Thus, a 2.5-fold increase in the percentage of functionality was achieved by randomly combining two siRNAs.
30 Further, over 84% of pools containing two siRNAs induced more than 80% gene silencing.

More preferably, the kit is comprised of at least three siRNAs, wherein one strand of each siRNA comprises a sequence that is substantially similar to a sequence

of the target mRNA and the other strand comprises a sequence that is substantially complementary to the region of the target mRNA. As with the kit that comprises at least two siRNAs, more preferably one strand will comprise a sequence that is identical to a sequence that is contained in the mRNA and another strand that is 100% complementary to a sequence that is contained in the mRNA. During experiments, when three siRNAs were combined together, 60% of the pools induced more than 95% gene silencing and 92% of the pools induced more than 80% gene silencing.

Further, even more preferably, the kit is comprised of at least four siRNAs, wherein one strand of each siRNA comprises a sequence that is substantially similar to a region of the sequence of the target mRNA, and the other strand comprises a sequence that is substantially complementary to the region of the target mRNA. As with the kit or pool that comprises at least two siRNAs, more preferably one strand of each of the siRNA duplexes will comprise a sequence that is identical to a sequence that is contained in the mRNA, and another strand that is 100% complementary to a sequence that is contained in the mRNA.

Additionally, kits and pools with at least five, at least six, and at least seven siRNAs may also be useful with the present invention. For example, pools of five siRNA induced 95% gene silencing with 77% probability and 80% silencing with 98.8% probability. Thus, pooling of siRNAs together can result in the creation of a target-specific silencing reagent with almost a 99% probability of being functional. The fact that such high levels of success are achievable using such pools of siRNA, enables one to dispense with costly and time-consuming target-specific validation procedures.

For this embodiment, as well as the other aforementioned embodiments, each of the siRNAs within a pool will preferably comprise between 18 and 30 base pairs, more preferably between 18 and 25 base pairs, and most preferably 19 base pairs. Within each siRNA, preferably at least 18 contiguous bases of the antisense strand will be 100% complementary to the target mRNA. More preferably, at least 19 contiguous bases of the antisense strand will be 100% complementary to the target mRNA. Additionally, there may be overhangs on either the sense strand or the antisense strand, and these overhangs may be at either the 5' end or the 3' end of

either of the strands, for example there may be one or more overhangs of 1-6 bases. When overhangs are present, they are not included in the calculation of the number of base pairs. The two nucleotide 3' overhangs mimic natural siRNAs and are commonly used but are not essential. Preferably, the overhangs should consist of two nucleotides, most often dTdT or UU at the 3' end of the sense and antisense strand that are not complementary to the target sequence. The siRNAs may be produced by any method that is now known or that comes to be known for synthesizing double stranded RNA that one skilled in the art would appreciate would be useful in the present invention. Preferably, the siRNAs will be produced by Dharmacon's proprietary ACE® technology. However, other methods for synthesizing siRNAs are well known to persons skilled in the art and include, but are not limited to, any chemical synthesis of RNA oligonucleotides, ligation of shorter oligonucleotides, *in vitro* transcription of RNA oligonucleotides, the use of vectors for expression within cells, recombinant Dicer products and PCR products.

The siRNA duplexes within the aforementioned pools of siRNAs may correspond to overlapping sequences within a particular mRNA, or non-overlapping sequences of the mRNA. However, preferably they correspond to non-overlapping sequences. Further, each siRNA may be selected randomly, or one or more of the siRNA may be selected according to the criteria discussed above for maximizing the effectiveness of siRNA.

Included in the definition of siRNAs are siRNAs that contain substituted and/or labeled nucleotides that may, for example, be labeled by radioactivity, fluorescence or mass. The most common substitutions are at the 2' position of the ribose sugar, where moieties such as H (hydrogen) F, NH₃, OCH₃ and other O- alkyl, alkenyl, alkynyl, and orthoesters, may be substituted, or in the phosphorous backbone, where sulfur, amines or hydrocarbons may be substituted for the bridging of non-bridging atoms in the phosphodiester bond. Examples of modified siRNAs are explained more fully in commonly assigned U.S. Patent Application Ser. No. 10/613,077, filed July 1, 2003, which is incorporated by reference herein.

Additionally, as noted above, the cell type into which the siRNA is introduced may affect the ability of the siRNA to enter the cell; however, it does not appear to

affect the ability of the siRNA to function once it enters the cell. Methods for introducing double-stranded RNA into various cell types are well known to persons skilled in the art.

5 As persons skilled in the art are aware, in certain species, the presence of proteins such as RdRP, the RNA-dependent RNA polymerase, may catalytically enhance the activity of the siRNA. For example, RdRP propagates the RNAi effect in *C. elegans* and other non-mammalian organisms. In fact, in organisms that contain these proteins, the siRNA may be inherited. Two other proteins that are well studied
10 and known to be a part of the machinery are members of the Argonaute family and Dicer, as well as their homologues. There is also initial evidence that the RISC complex might be associated with the ribosome so the more efficiently translated mRNAs will be more susceptible to silencing than others.

15 Another very important factor in the efficacy of siRNA is mRNA localization. In general, only cytoplasmic mRNAs are considered to be accessible to RNAi to any appreciable degree. However, appropriately designed siRNAs, for example, siRNAs modified with internucleotide linkages, may be able to cause silencing by acting in the nucleus. Examples of these types of modifications are described in commonly
20 assigned U.S. Patent Application Serial Nos. 10/431,027 and 10/613,077, each of which is incorporated by reference herein.

 As described above, even when one selects at least two siRNAs at random, the effectiveness of the two may be greater than one would predict based on the
25 effectiveness of two individual siRNAs. This additive or synergistic effect is particularly noticeable as one increases to at least three siRNAs, and even more noticeable as one moves to at least four siRNAs. Surprisingly, the pooling of the non-functional and semi-functional siRNAs, particularly more than five siRNAs, can lead to a silencing mixture that is as effective if not more effective than any one particular
30 functional siRNA.

Within the kit of the present invention, preferably each siRNA will be present in a concentration of between 0.001 and 200 μM , more preferably between 0.01 and 200 nM, and most preferably between 0.1 and 10 nM.

5 In addition to preferably comprising at least four or five siRNAs, the kit of the present invention will also preferably comprise a buffer to keep the siRNA duplex stable. Persons skilled in the art are aware of buffers suitable for keeping siRNA stable. For example, the buffer may be comprised of 100 mM KCl, 30 mM HEPES-
10 pH 7.5, and 1 mM MgCl_2 . Alternatively, kits might contain complementary strands that contain any one of a number of chemical modifications (e.g. a 2'-O-ACE) that protect the agents from degradation by nucleases. In this instance, the user may (or may not) remove the modifying protective group (e.g. deprotect) before annealing the two complementary strands together.

15 By way of example, the kit may be organized such that pools of siRNA duplexes are provided on an array or microarray of wells or drops for a particular gene set or for unrelated genes. The array may, for example, be in 96 wells, 384 wells or 1284 wells arrayed in a plastic plate or on a glass slide using techniques now
20 known or that come to be known to persons skilled in the art. Within an array, preferably there will be controls such as functional anti-lamin A/C, cyclophilin and two siRNA duplexes that are not specific to the gene of interest.

25 In order to ensure stability of the siRNA pools prior to usage, they may be retained in lyophilized form at minus twenty degrees (-20°C) until they are ready for use. Prior to usage, they should be resuspended; however, even once resuspended, for example, in the aforementioned buffer, they should be kept at minus twenty degrees, (-20°C) until used. The aforementioned buffer, prior to use, may be stored at
30 approximately 4°C or room temperature. Effective temperatures at which to conduct transfections are well known to persons skilled in the art and include for example, room temperature.

The kit may be applied either *in vivo* or *in vitro*. Preferably, the siRNA of the pools or kits is applied to a cell through transfection, employing standard transfection

protocols. These methods are well known to persons skilled in the art and include the use of lipid-based carriers, electroporation, cationic carriers, and microinjection.

Further, one could apply the present invention by synthesizing equivalent DNA sequences (either as two separate, complementary strands, or as hairpin molecules)

5 instead of siRNA sequences and introducing them into cells through vectors. Once in the cells, the cloned DNA could be transcribed, thereby forcing the cells to generate the siRNA. Examples of vectors suitable for use with the present application include but are not limited to the standard transient expression vectors, adenoviruses, retroviruses, lentivirus-based vectors, as well as other traditional expression vectors.

10 Any vector that has an adequate siRNA expression and procession module may be used. Furthermore, certain chemical modifications to siRNAs, including but not limited to conjugations to other molecules, may be used to facilitate delivery. For certain applications it may be preferable to deliver molecules without transfection by simply formulating in a physiological acceptable solution.

15 This embodiment may be used in connection with any of the aforementioned embodiments. Accordingly, the sequences within any pool may be selected by rational design.

20 **Multigene Silencing**

In addition to developing kits that contain multiple siRNA directed against a single gene, another embodiment includes the use of multiple siRNA targeting multiple genes. Multiple genes may be targeted through the use of high- or hyper-

25 functional siRNA. High- or hyper- functional siRNA that exhibit increased potency, require lower concentrations to induce desired phenotypic (and thus therapeutic) effects. This circumvents RISC saturation. It therefore reasons that if lower concentrations of a single siRNA are needed for knockout or knockdown expression of one gene, then the remaining (uncomplexed) RISC will be free and available to

30 interact with siRNA directed against two, three, four, or more, genes. Thus in this embodiment, the authors describe the use of highly functional or hyper-functional siRNA to knock out three separate genes. More preferably, such reagents could be combined to knockout four distinct genes. Even more preferably, highly functional or hyperfunctional siRNA could be used to knock out five distinct genes. Most

preferably, siRNA of this type could be used to knockout or knockdown the expression of six or more genes.

Hyperfunctional siRNA

5 The term hyperfunctional siRNA (hf-siRNA) describes a subset of the siRNA population that induces RNAi in cells at low- or sub-nanomolar concentrations for extended periods of time. These traits, heightened potency and extended longevity of the RNAi phenotype, are highly attractive from a therapeutic standpoint. Agents having higher potency require lesser amounts of the molecule to achieve the desired physiological response, thus reducing the probability of side effects due to “off-target” interference. In addition to the potential therapeutic benefits associated with hyperfunctional siRNA, hf-siRNA are also desirable from an economic position. Hyperfunctional siRNA may cost less on a per-treatment basis, thus reducing the overall expenditures to both the manufacturer and the consumer.

10 Identification of hyperfunctional siRNA involves multiple steps that are designed to examine an individual siRNA agent’s concentration- and/or longevity-profiles. In one non-limiting example, a population of siRNA directed against a single gene are first analyzed using the previously described algorithm (Formula VIII). Individual siRNA are then introduced into a test cell line and assessed for the ability to degrade the target mRNA. It is important to note that when performing this step it is not necessary to test all of the siRNA. Instead, it is sufficient to test only those siRNA having the highest SMARTscores™ (*i.e.* SMARTscore™ > -10). Subsequently, the gene silencing data is plotted against the SMARTscores™ (see **Figure 9**). SiRNA that (1) induce a high degree of gene silencing (*i.e.* they induce greater than 80% gene knockdown) and (2) have superior SMARTscores™ (*i.e.* a SMARTscore™ of > -10, suggesting a desirable average internal stability profile) are selected for further investigations designed to better understand the molecule’s potency and longevity. In one, non-limiting study dedicated to understanding a molecule’s potency, an siRNA is introduced into one (or more) cell types in increasingly diminishing concentrations (*e.g.* 3.0 → 0.3 nM). Subsequently, the level of gene silencing induced by each concentration is examined and siRNA that exhibit hyperfunctional potency (*i.e.* those that induce 80% silencing or greater at *e.g.* picomolar concentrations) are identified. In a second study, the longevity profiles of

15
20
25
30

siRNA having high (>-10) SMARTscores™ and greater than 80% silencing are examined. In one non-limiting example of how this is achieved, siRNA are introduced into a test cell line and the levels of RNAi are measured over an extended period of time (*e.g.* 24-168 hrs). SiRNAs that exhibit strong RNA interference patterns (*i.e.* $>80\%$ interference) for periods of time greater than, *e.g.*, 120 hours are thus identified. Studies similar to those described above can be performed on any and all of the $>10^6$ siRNA included in this document to further define the most functional molecule for any given gene. Molecules possessing one or both properties (extended longevity and heightened potency) are labeled “hyperfunctional siRNA,” and earmarked as candidates for future therapeutic studies.

While the example(s) given above describe one means by which hyperfunctional siRNA can be isolated, neither the assays themselves nor the selection parameters used are rigid and can vary with each family of siRNA. Families of siRNA include siRNAs directed against a single gene, or directed against a related family of genes.

The highest quality siRNA achievable for any given gene may vary considerably. Thus, for example, in the case of one gene (gene X), rigorous studies such as those described above may enable the identification of an siRNA that, at picomolar concentrations, induces $99^{+}\%$ silencing for a period of 10 days. Yet identical studies of a second gene (gene Y) may yield an siRNA that at high nanomolar concentrations (*e.g.* 100nM) induces only 75% silencing for a period of 2 days. Both molecules represent the very optimum siRNA for their respective gene targets and therefore are designated “hyperfunctional.” Yet due to a variety of factors including but not limited to target concentration, siRNA stability, cell type, off-target interference, and others, equivalent levels of potency and longevity are not achievable. Thus, for these reasons, the parameters described in the before mentioned assays, can vary. While the initial screen selected siRNA that had SMARTscores™ above -10 and a gene silencing capability of greater than 80%, selections that have stronger (or weaker) parameters can be implemented. Similarly, in the subsequent studies designed to identify molecules with high potency and longevity, the desired cutoff criteria (*i.e.* the lowest concentration that induces a desirable level of interference, or the longest period of time that interference can be observed) can vary.

The experimentation subsequent to application of the rational criteria of this application is significantly reduced where one is trying to obtain a suitable hyperfunctional siRNA for, for example, therapeutic use. When, for example, the additional experimentation of the type described herein is applied by one skilled in the art with this disclosure in hand, a hyperfunctional siRNA is readily identified.

The siRNA may be introduced into a cell by any method that is now known or that comes to be known and that from reading this disclosure, persons skilled in the art would determine would be useful in connection with the present invention in enabling siRNA to cross the cellular membrane. These methods include, but are not limited to, any manner of transfection, such as for example transfection employing DEAE-Dextran, calcium phosphate, cationic lipids/liposomes, micelles, manipulation of pressure, microinjection, electroporation, immunoporation, use of vectors such as viruses, plasmids, cosmids, bacteriophages, cell fusions, and coupling of the polynucleotides to specific conjugates or ligands such as antibodies, antigens, or receptors, passive introduction, adding moieties to the siRNA that facilitate its uptake, and the like.

Having described the invention with a degree of particularity, examples will now be provided. These examples are not intended to and should not be construed to limit the scope of the claims in any way.

Examples

25 **General Techniques and Nomenclatures**

siRNA nomenclature. All siRNA duplexes are referred to by sense strand. The first nucleotide of the 5'-end of the sense strand is position 1, which corresponds to position 19 of the antisense strand for a 19-mer. In most cases, to compare results from different experiments, silencing was determined by measuring specific transcript mRNA levels or enzymatic activity associated with specific transcript levels, 24 hours post-transfection, with siRNA concentrations held constant at 100 nM. For all experiments, unless otherwise specified transfection efficiency was ensured to be over 95%, and no detectable cellular toxicity was observed. The following system of nomenclature was used to compare and report siRNA-silencing functionality: "F"

followed by the degree of minimal knockdown. For example, F50 signifies at least 50% knockdown, F80 means at least 80%, and so forth. For this study, all sub-F50 siRNAs were considered non-functional.

5 Cell culture and transfection. 96-well plates are coated with 50 μ l of 50 mg/ml poly-L-lysine (Sigma) for 1 hr, and then washed 3X with distilled water before being dried for 20 min. HEK293 cells or HEK293Lucs or any other cell type of interest are released from their solid support by trypsinization, diluted to 3.5×10^5 cells/ml, followed by the addition of 100 μ L of cells/well. Plates are then incubated overnight
10 at 37° C, 5% CO₂. Transfection procedures can vary widely depending on the cell type and transfection reagents. In one non-limiting example, a transfection mixture consisting of 2 mL Opti-MEM I (Gibco-BRL), 80 μ l Lipofectamine 2000 (Invitrogen), 15 μ L SUPERNasin at 20 U/ μ l (Ambion), and 1.5 μ l of reporter gene plasmid at 1 μ g/ μ l is prepared in 5-ml polystyrene round bottom tubes. 100 μ l of
15 transfection reagent is then combined with 100 μ l of siRNAs in polystyrene deep-well titer plates (Beckman) and incubated for 20 to 30 min at room temp. 550 μ l of Opti-MEM is then added to each well to bring the final siRNA concentration to 100 nM. Plates are then sealed with parafilm and mixed. Media is removed from HEK293 cells and replaced with 95 μ l of transfection mixture. Cells are incubated overnight at
20 37° C, 5% CO₂.

Quantification of gene knockdown. A variety of quantification procedures can be used to measure the level of silencing induced by siRNA or siRNA pools. In one non-limiting example: to measure mRNA levels 24 hrs post-transfection, QuantiGene
25 branched-DNA (bDNA) kits (Bayer) (Wang, *et al*, *Regulation of insulin preRNA splicing by glucose*. Proc Natl Acad Sci 1997, 94:4360.) are used according to manufacturer instructions. To measure luciferase activity, media is removed from HEK293 cells 24 hrs post-transfection, and 50 μ l of Steady-GLO reagent (Promega) is added. After 5 min, plates are analyzed on a plate reader.

30

Example I. Sequences Used to Develop the Algorithm.

Anti-Firefly and anti-Cyclophilin siRNAs panels (Figure 5a, b) sorted according to using Formula VIII predicted values. All siRNAs scoring more than 0

(formula VIII) and more than 20 (formula IX) are fully functional. All ninety sequences for each gene (and DBI) appear below in **Table III**.

5

TABLE III

Cyclo	1	SEQ. ID 0032	GUUCCAAAAACAGUGGAUA
Cyclo	2	SEQ. ID 0033	UCCAAAAACAGUGGAUAAU
Cyclo	3	SEQ. ID 0034	CAAAAACAGUGGAUAAUUU
Cyclo	4	SEQ. ID 0035	AAAACAGUGGAUAAUUUUG
Cyclo	5	SEQ. ID 0036	AACAGUGGAUAAUUUUGUG
Cyclo	6	SEQ. ID 0037	CAGUGGAUAAUUUUGUGGC
Cyclo	7	SEQ. ID 0038	GUGGAUAAUUUUGUGGCCU
Cyclo	8	SEQ. ID 0039	GAUAAUUUUGUGGCCUUA
Cyclo	9	SEQ. ID 0040	AUAAUUUUGUGGCCUAGC
Cyclo	10	SEQ. ID 0041	AAUUUUGUGGCCUAGCUA
Cyclo	11	SEQ. ID 0042	UUUUGUGGCCUAGCUACA
Cyclo	12	SEQ. ID 0043	UUGUGGCCUAGCUACAGG
Cyclo	13	SEQ. ID 0044	GUGGCCUAGCUACAGGAG
Cyclo	14	SEQ. ID 0045	GGCCUAGCUACAGGAGAG
Cyclo	15	SEQ. ID 0046	CCUAGCUACAGGAGAGAA
Cyclo	16	SEQ. ID 0047	UUAGCUACAGGAGAGAAAG
Cyclo	17	SEQ. ID 0048	AGCUACAGGAGAGAAAGGA
Cyclo	18	SEQ. ID 0049	CUACAGGAGAGAAAGGAUU
Cyclo	19	SEQ. ID 0050	ACAGGAGAGAAAGGAUUUG
Cyclo	20	SEQ. ID 0051	AGGAGAGAAAGGAUUUGGC
Cyclo	21	SEQ. ID 0052	GAGAGAAAGGAUUUGGCUA
Cyclo	22	SEQ. ID 0053	GAGAAAGGAUUUGGCUACA
Cyclo	23	SEQ. ID 0054	GAAAGGAUUUGGCUACAAA
Cyclo	24	SEQ. ID 0055	AAGGAUUUGGCUACAAAAA
Cyclo	25	SEQ. ID 0056	GGAUUUGGCUACAAAAACA
Cyclo	26	SEQ. ID 0057	AUUUGGCUACAAAAACAGC
Cyclo	27	SEQ. ID 0058	UUGGCUACAAAAACAGCAA
Cyclo	28	SEQ. ID 0059	GGCUACAAAAACAGCAAUU
Cyclo	29	SEQ. ID 0060	CUACAAAAACAGCAAUUUC
Cyclo	30	SEQ. ID 0061	ACAAAAACAGCAAUUCCA
Cyclo	31	SEQ. ID 0062	AAAAACAGCAAUUCCAUC
Cyclo	32	SEQ. ID 0063	AAACAGCAAUUCCAUCGU
Cyclo	33	SEQ. ID 0064	ACAGCAAUUCCAUCGUGU
Cyclo	34	SEQ. ID 0065	AGCAAUUCCAUCGUGUAA

Cyclo	35	SEQ. ID 0066	CAAAUUCCAUCGUGUAAUUC
Cyclo	36	SEQ. ID 0067	AAUUCCAUCGUGUAAUCAA
Cyclo	37	SEQ. ID 0068	UUCCAUCGUGUAAUCAAGG
Cyclo	38	SEQ. ID 0069	CCAUCGUGUAAUCAAGGAC
Cyclo	39	SEQ. ID 0070	AUCGUGUAAUCAAGGACUU
Cyclo	40	SEQ. ID 0071	CGUGUAAUCAAGGACUUCA
Cyclo	41	SEQ. ID 0072	UGUAAUCAAGGACUUCAUG
Cyclo	42	SEQ. ID 0073	UAAUCAAGGACUUCAUGAU
Cyclo	43	SEQ. ID 0074	AUCAAGGACUUCAUGAUCC
Cyclo	44	SEQ. ID 0075	CAAGGACUUCAUGAUCCAG
Cyclo	45	SEQ. ID 0076	AGGACUUCAUGAUCCAGGG
Cyclo	46	SEQ. ID 0077	GACUUCAUGAUCCAGGGCG
Cyclo	47	SEQ. ID 0078	CUUCAUGAUCCAGGGCGGA
Cyclo	48	SEQ. ID 0079	UCAUGAUCCAGGGCGGAGA
Cyclo	49	SEQ. ID 0080	AUGAUCCAGGGCGGAGACU
Cyclo	50	SEQ. ID 0081	GAUCCAGGGCGGAGACUUC
Cyclo	51	SEQ. ID 0082	UCCAGGGCGGAGACUUCAC
Cyclo	52	SEQ. ID 0083	CAGGGCGGAGACUUCACCA
Cyclo	53	SEQ. ID 0084	GGGCGGAGACUUCACCAGG
Cyclo	54	SEQ. ID 0085	GCGGAGACUUCACCAGGGG
Cyclo	55	SEQ. ID 0086	GGAGACUUCACCAGGGGAG
Cyclo	56	SEQ. ID 0087	AGACUUCACCAGGGGAGAU
Cyclo	57	SEQ. ID 0088	ACUUCACCAGGGGAGAUUGG
Cyclo	58	SEQ. ID 0089	UUCACCAGGGGAGAUUGCA
Cyclo	59	SEQ. ID 0090	CACCAGGGGAGAUUGCACA
Cyclo	60	SEQ. ID 0091	CCAGGGGAGAUUGGCACAGG
Cyclo	61	SEQ. ID 0092	AGGGGAGAUUGGCACAGGAG
Cyclo	62	SEQ. ID 0093	GGGAGAUUGGCACAGGAGGA
Cyclo	63	SEQ. ID 0094	GAGAUGGCACAGGAGGAAA
Cyclo	64	SEQ. ID 0095	GAUGGCACAGGAGGAAAGA
Cyclo	65	SEQ. ID 0094	UGGCACAGGAGGAAAGAGC
Cyclo	66	SEQ. ID 0096	GCACAGGAGGAAAGAGCAU
Cyclo	67	SEQ. ID 0097	ACAGGAGGAAAGAGCAUCU
Cyclo	68	SEQ. ID 0098	AGGAGGAAAGAGCAUCUAC
Cyclo	69	SEQ. ID 0099	GAGGAAAGAGCAUCUACGG
Cyclo	70	SEQ. ID 0100	GGAAAGAGCAUCUACGGUG
Cyclo	71	SEQ. ID 0101	AAAGAGCAUCUACGGUGAG
Cyclo	72	SEQ. ID 0102	AGAGCAUCUACGGUGAGCG
Cyclo	73	SEQ. ID 0103	AGCAUCUACGGUGAGCGCU

Cyclo	74	SEQ. ID 0104	CAUCUACGGUGAGCGCUUC
Cyclo	75	SEQ. ID 0105	UCUACGGUGAGCGCUUCCC
Cyclo	76	SEQ. ID 0106	UACGGUGAGCGCUUCCCCG
Cyclo	77	SEQ. ID 0107	CGGUGAGCGCUUCCCCGAU
Cyclo	78	SEQ. ID 0108	GUGAGCGCUUCCCCGAUGA
Cyclo	79	SEQ. ID 0109	GAGCGCUUCCCCGAUGAGA
Cyclo	80	SEQ. ID 0110	GCGCUUCCCCGAUGAGAAC
Cyclo	81	SEQ. ID 0111	GCUUCCCCGAUGAGAACUU
Cyclo	82	SEQ. ID 0112	UUCCCCGAUGAGAACUUA
Cyclo	83	SEQ. ID 0113	CCCCGAUGAGAACUUCAAA
Cyclo	84	SEQ. ID 0114	CCGAUGAGAACUUCAAACU
Cyclo	85	SEQ. ID 0115	GAUGAGAACUUCAAACUGA
Cyclo	86	SEQ. ID 0116	UGAGAACUUCAAACUGAAG
Cyclo	87	SEQ. ID 0117	AGAACUUCAAACUGAAGCA
Cyclo	88	SEQ. ID 0118	AACUUCAAACUGAAGCACU
Cyclo	89	SEQ. ID 0119	CUUCAAAACUGAAGCACUAC
Cyclo	90	SEQ. ID 0120	UCAAACUGAAGCACUACGG
DB	1	SEQ. ID 0121	ACGGGCAAGGCCAAGUGGG
DB	2	SEQ. ID 0122	CGGGCAAGGCCAAGUGGGA
DB	3	SEQ. ID 0123	GGGCAAGGCCAAGUGGGAU
DB	4	SEQ. ID 0124	GGCAAGGCCAAGUGGGAUG
DB	5	SEQ. ID 0125	GCAAGGCCAAGUGGGAUGC
DB	6	SEQ. ID 0126	CAAGGCCAAGUGGGAUGCC
DB	7	SEQ. ID 0127	AAGGCCAAGUGGGAUGCCU
DB	8	SEQ. ID 0128	AGGCCAAGUGGGAUGCCUG
DB	9	SEQ. ID 0129	GGCCAAGUGGGAUGCCUGG
DB	10	SEQ. ID 0130	GCCAAGUGGGAUGCCUGGA
DB	11	SEQ. ID 0131	CCAAGUGGGAUGCCUGGAA
DB	12	SEQ. ID 0132	CAAGUGGGAUGCCUGGAAU
DB	13	SEQ. ID 0133	AAGUGGGAUGCCUGGAAUG
DB	14	SEQ. ID 0134	AGUGGGAUGCCUGGAAUGA
DB	15	SEQ. ID 0135	GUGGGAUGCCUGGAAUGAG
DB	16	SEQ. ID 0136	UGGGAUGCCUGGAAUGAGC
DB	17	SEQ. ID 0137	GGGAUGCCUGGAAUGAGCU
DB	18	SEQ. ID 0138	GGAUGCCUGGAAUGAGCUG
DB	19	SEQ. ID 0139	GAUGCCUGGAAUGAGCUGA
DB	20	SEQ. ID 0140	AUGCCUGGAAUGAGCUGAA
DB	21	SEQ. ID 0141	UGCCUGGAAUGAGCUGAAA
DB	22	SEQ. ID 0142	GCCUGGAAUGAGCUGAAAG
DB	23	SEQ. ID 0143	CCUGGAAUGAGCUGAAAGG

DB	24	SEQ. ID 0144	CUGGAAUGAGCUGAAAGGG
DB	25	SEQ. ID 0145	UGGAAUGAGCUGAAAGGGA
DB	26	SEQ. ID 0146	GGAAUGAGCUGAAAGGGAC
DB	27	SEQ. ID 0147	GAAUGAGCUGAAAGGGACU
DB	28	SEQ. ID 0148	AAUGAGCUGAAAGGGACUU
DB	29	SEQ. ID 0149	AUGAGCUGAAAGGGACUUC
DB	30	SEQ. ID 0150	UGAGCUGAAAGGGACUCC
DB	31	SEQ. ID 0151	GAGCUGAAAGGGACUCCA
DB	32	SEQ. ID 0152	AGCUGAAAGGGACUCCAA
DB	33	SEQ. ID 0153	GCUGAAAGGGACUCCAAG
DB	34	SEQ. ID 0154	CUGAAAGGGACUCCAAGG
DB	35	SEQ. ID 0155	UGAAAGGGACUCCAAGGA
DB	36	SEQ. ID 0156	GAAAGGGACUCCAAGGAA
DB	37	SEQ. ID 0157	AAAGGGACUCCAAGGAAG
DB	38	SEQ. ID 0158	AAGGGACUCCAAGGAAGA
DB	39	SEQ. ID 0159	AGGGACUCCAAGGAAGAU
DB	40	SEQ. ID 0160	GGGACUCCAAGGAAGAUG
DB	41	SEQ. ID 0161	GGACUCCAAGGAAGAUGC
DB	42	SEQ. ID 0162	GACUCCAAGGAAGAUGCC
DB	43	SEQ. ID 0163	ACUCCAAGGAAGAUGCCA
DB	44	SEQ. ID 0164	CUCCAAGGAAGAUGCCAU
DB	45	SEQ. ID 0165	UCCAAGGAAGAUGCCAUG
DB	46	SEQ. ID 0166	UCCAAGGAAGAUGCCAUGA
DB	47	SEQ. ID 0167	CCAAGGAAGAUGCCAUGAA
DB	48	SEQ. ID 0168	CAAGGAAGAUGCCAUGAAA
DB	49	SEQ. ID 0169	AAGGAAGAUGCCAUGAAAG
DB	50	SEQ. ID 0170	AGGAAGAUGCCAUGAAAGC
DB	51	SEQ. ID 0171	GGAAGAUGCCAUGAAAGCU
DB	52	SEQ. ID 0172	GAAGAUGCCAUGAAAGCUU
DB	53	SEQ. ID 0173	AAGAUGCCAUGAAAGCUUA
DB	54	SEQ. ID 0174	AGAUGCCAUGAAAGCUUAC
DB	55	SEQ. ID 0175	GAUGCCAUGAAAGCUUACA
DB	56	SEQ. ID 0176	AUGCCAUGAAAGCUUACAU
DB	57	SEQ. ID 0177	UGCCAUGAAAGCUUACAUC
DB	58	SEQ. ID 0178	GCCAUGAAAGCUUACAUCA
DB	59	SEQ. ID 0179	CCAUGAAAGCUUACAUCAA
DB	60	SEQ. ID 0180	CAUGAAAGCUUACAUCAAC
DB	61	SEQ. ID 0181	AUGAAAGCUUACAUCAACA
DB	62	SEQ. ID 0182	UGAAAGCUUACAUCAACAA
DB	63	SEQ. ID 0183	GAAAGCUUACAUCAACAAA

DB	64	SEQ. ID 0184	AAAGCUUACAUCAACAAAG
DB	65	SEQ. ID 0185	AAGCUUACAUCAACAAAGU
DB	66	SEQ. ID 0186	AGCUUACAUCAACAAAGUA
DB	67	SEQ. ID 0187	GCUUACAUCAACAAAGUAG
DB	68	SEQ. ID 0188	CUUACAUCAACAAAGUAGA
DB	69	SEQ. ID 0189	UUACAUCAACAAAGUAGAA
DB	70	SEQ. ID 0190	UACAUCAACAAAGUAGAAG
DB	71	SEQ. ID 0191	ACAUCAACAAAGUAGAAGA
DB	72	SEQ. ID 0192	CAUCAACAAAGUAGAAGAG
DB	73	SEQ. ID 0193	AUCAACAAAGUAGAAGAGC
DB	74	SEQ. ID 0194	UCAACAAAGUAGAAGAGCU
DB	75	SEQ. ID 0195	CAACAAAGUAGAAGAGCUA
DB	76	SEQ. ID 0196	AACAAAGUAGAAGAGCUAA
DB	77	SEQ. ID 0197	ACAAAGUAGAAGAGCUAAA
DB	78	SEQ. ID 0198	CAAAGUAGAAGAGCUAAAG
DB	79	SEQ. ID 0199	AAAGUAGAAGAGCUAAAGA
DB	80	SEQ. ID 0200	AAGUAGAAGAGCUAAAGAA
DB	81	SEQ. ID 0201	AGUAGAAGAGCUAAAGAAA
DB	82	SEQ. ID 0202	GUAGAAGAGCUAAAGAAAA
DB	83	SEQ. ID 0203	UAGAAGAGCUAAAGAAAAA
DB	84	SEQ. ID 0204	AGAAGAGCUAAAGAAAAAA
DB	85	SEQ. ID 0205	GAAGAGCUAAAGAAAAAAU
DB	86	SEQ. ID 0206	AAGAGCUAAAGAAAAAAUA
DB	87	SEQ. ID 0207	AGAGCUAAAGAAAAAAUAC
DB	88	SEQ. ID 0208	GAGCUAAAGAAAAAAUACG
DB	89	SEQ. ID 0209	AGCUAAAGAAAAAAUACGG
DB	90	SEQ. ID 0210	GCUAAAGAAAAAAUACGGG
Luc	1	SEQ. ID 0211	AUCCUCAUAAAGGCCAAGA
Luc	2	SEQ. ID 0212	AGAUCCUCAUAAAGGCCAA
Luc	3	SEQ. ID 0213	AGAGAUCCUCAUAAAGGCC
Luc	4	SEQ. ID 0214	AGAGAGAUCCUCAUAAAGG
Luc	5	SEQ. ID 0215	UCAGAGAGAUCCUCAUAAA
Luc	6	SEQ. ID 0216	AAUCAGAGAGAUCCUCAUA
Luc	7	SEQ. ID 0217	AAAAUCAGAGAGAUCCUCA
Luc	8	SEQ. ID 0218	GAAAAUCAGAGAGAUCCU
Luc	9	SEQ. ID 0219	AAGAAAAUCAGAGAGAUCC
Luc	10	SEQ. ID 0220	GCAAGAAAAUCAGAGAGAG
Luc	11	SEQ. ID 0221	ACGCAAGAAAAUCAGAGAG
Luc	12	SEQ. ID 0222	CGACGCAAGAAAAUCAGAG
Luc	13	SEQ. ID 0223	CUCGACGCAAGAAAAUCA

Luc	14	SEQ. ID 0224	AACUCGACGCAAGAAAAAU
Luc	15	SEQ. ID 0225	AAAACUCGACGCAAGAAAA
Luc	16	SEQ. ID 0226	GGAAAACUCGACGCAAGAA
Luc	17	SEQ. ID 0227	CCGGAAAACUCGACGCAAG
Luc	18	SEQ. ID 0228	UACCGGAAAACUCGACGCA
Luc	19	SEQ. ID 0229	CUUACCGGAAAACUCGACG
Luc	20	SEQ. ID 0230	GUCUUACCGGAAAACUCGA
Luc	21	SEQ. ID 0231	AGGUCUUACCGGAAAACUC
Luc	22	SEQ. ID 0232	AAAGGUCUUACCGGAAAAC
Luc	23	SEQ. ID 0233	CGAAAGGUCUUACCGGAAA
Luc	24	SEQ. ID 0234	ACCGAAAGGUCUUACCGGA
Luc	25	SEQ. ID 0235	GUACCGAAAGGUCUUACCG
Luc	26	SEQ. ID 0236	AAGUACCGAAAGGUCUUAC
Luc	27	SEQ. ID 0237	CGAAGUACCGAAAGGUCUU
Luc	28	SEQ. ID 0238	GACGAAGUACCGAAAGGUC
Luc	29	SEQ. ID 0239	UGGACGAAGUACCGAAAGG
Luc	30	SEQ. ID 0240	UGUGGACGAAGUACCGAAA
Luc	31	SEQ. ID 0241	UUUGUGGACGAAGUACCGA
Luc	32	SEQ. ID 0242	UGUUUGUGGACGAAGUACC
Luc	33	SEQ. ID 0243	UGUGUUUGUGGACGAAGUA
Luc	34	SEQ. ID 0244	GUUGUGUUUGUGGACGAAG
Luc	35	SEQ. ID 0245	GAGUUGUGUUUGUGGACGA
Luc	36	SEQ. ID 0246	AGGAGUUGUGUUUGUGGAC
Luc	37	SEQ. ID 0247	GGAGGAGUUGUGUUUGUGG
Luc	38	SEQ. ID 0248	GCGGAGGAGUUGUGUUUGU
Luc	39	SEQ. ID 0249	GCGCGGAGGAGUUGUGUUU
Luc	40	SEQ. ID 0250	UUGCGCGGAGGAGUUGUGU
Luc	41	SEQ. ID 0251	AGUUGCGCGGAGGAGUUGU
Luc	42	SEQ. ID 0252	AAAGUUGCGCGGAGGAGUU
Luc	43	SEQ. ID 0253	AAAAAGUUGCGCGGAGGAG
Luc	44	SEQ. ID 0254	CGAAAAAGUUGCGCGGAGG
Luc	45	SEQ. ID 0255	CGCGAAAAAGUUGCGCGGA
Luc	46	SEQ. ID 0256	ACCGCGAAAAAGUUGCGCG
Luc	47	SEQ. ID 0257	CAACCGCGAAAAAGUUGCG
Luc	48	SEQ. ID 0258	AACAACCGCGAAAAAGUUG
Luc	49	SEQ. ID 0259	GUAACAACCGCGAAAAAGU
Luc	50	SEQ. ID 0260	AAGUAACAACCGCGAAAAA
Luc	51	SEQ. ID 0261	UCAAGUAACAACCGCGAAA
Luc	52	SEQ. ID 0262	AGUCAAGUAACAACCGCGA
Luc	53	SEQ. ID 0263	CCAGUCAAGUAACAACCGC

Luc	54	SEQ. ID 0264	CGCCAGUCAAGUAACAACC
Luc	55	SEQ. ID 0265	GUCGCCAGUCAAGUAACAA
Luc	56	SEQ. ID 0266	ACGUCGCCAGUCAAGUAAC
Luc	57	SEQ. ID 0267	UUACGUCGCCAGUCAAGUA
Luc	58	SEQ. ID 0268	GAUUACGUCGCCAGUCAAG
Luc	59	SEQ. ID 0269	UGGAUUACGUCGCCAGUCA
Luc	60	SEQ. ID 0270	CGUGGAUUACGUCGCCAGU
Luc	61	SEQ. ID 0271	AUCGUGGAUUACGUCGCCA
Luc	62	SEQ. ID 0272	AGAUCGUGGAUUACGUCGC
Luc	63	SEQ. ID 0273	AGAGAUCGUGGAUUACGUC
Luc	64	SEQ. ID 0274	AAAGAGAUCGUGGAUUACG
Luc	65	SEQ. ID 0275	AAAAAGAGAUCGUGGAUUA
Luc	66	SEQ. ID 0276	GGAAAAAGAGAUCGUGGAU
Luc	67	SEQ. ID 0277	ACGGAAAAAGAGAUCGUGG
Luc	68	SEQ. ID 0278	UGACGGAAAAAGAGAUCGU
Luc	69	SEQ. ID 0279	GAUGACGGAAAAAGAGAUC
Luc	70	SEQ. ID 0280	ACGAUGACGGAAAAAGAGA
Luc	71	SEQ. ID 0281	AGACGAUGACGGAAAAAGA
Luc	72	SEQ. ID 0282	AAAGACGAUGACGGAAAAA
Luc	73	SEQ. ID 0283	GGAAAGACGAUGACGGAAA
Luc	74	SEQ. ID 0284	ACGGAAAGACGAUGACGGA
Luc	75	SEQ. ID 0285	GCACGGAAAGACGAUGACG
Luc	76	SEQ. ID 0286	GAGCACGGAAAGACGAUGA
Luc	77	SEQ. ID 0287	UGGAGCACGGAAAGACGAU
Luc	78	SEQ. ID 0288	UUUGGAGCACGGAAAGACG
Luc	79	SEQ. ID 0289	GUUUUGGAGCACGGAAAGA
Luc	80	SEQ. ID 0290	UUGUUUUGGAGCACGGAAA
Luc	81	SEQ. ID 0291	UGUUGUUUUGGAGCACGGA
Luc	82	SEQ. ID 0292	GUUGUUGUUUUGGAGCACG
Luc	83	SEQ. ID 0293	CCGUUGUUGUUUUGGAGCA
Luc	84	SEQ. ID 0294	CGCCGUUGUUGUUUUGGAG
Luc	85	SEQ. ID 0295	GCCGCCGUUGUUGUUUUGG
Luc	86	SEQ. ID 0296	CCGCCGCCGUUGUUGUUU
Luc	87	SEQ. ID 0297	UCCCGCCGCCGUUGUUGU
Luc	88	SEQ. ID 0298	CUUCCCGCCGCCGUUGUUG
Luc	89	SEQ. ID 0299	AACUCCCGCCGCCGUUGU
Luc	90	SEQ. ID 0300	UGAACUCCCGCCGCCGUU

Example II. Validation of the Algorithm using DBI, Luciferase, PLK, EGFR, and SEAP

The algorithm (Formula VIII) identified siRNAs for five genes, human DBI, firefly luciferase (fLuc), renilla luciferase (rLuc), human PLK, and human secreted alkaline phosphatase (SEAP). Four individual siRNAs were selected on the basis of their SMARTscores™ derived by analysis of their sequence using Formula VIII (all of the siRNAs would be selected with Formula IX as well) and analyzed for their ability to silence their targets' expression. In addition to the scoring, a BLAST search was conducted for each siRNA. To minimize the potential for off-target silencing effects, only those target sequences with more than three mismatches against unrelated sequences were selected. Semizarov, *et al*, *Specificity of short interfering RNA determined through gene expression signatures*. Proc. Natl. Acad. Sci. U.S.A. 2003, 100:6347. These duplexes were analyzed individually and in pools of 4 and compared with several siRNAs that were randomly selected. The functionality was measured a percentage of targeted gene knockdown as compared to controls. All siRNAs were transfected as described by the methods above at 100 nM concentration into HEK293 using Lipofectamine 2000. The level of the targeted gene expression was evaluated by B-DNA as described above and normalized to the non-specific control. Figure 10 shows that the siRNAs selected by the algorithm disclosed herein were significantly more potent than randomly selected siRNAs. The algorithm increased the chances of identifying an F50 siRNA from 48% to 91%, and an F80 siRNA from 13% to 57%. In addition, pools of SMART siRNA silence the selected target better than randomly selected pools (see Figure 10F).

Example III. Validation of the Algorithm Using Genes Involved in Clathrin-Dependent Endocytosis.

Components of clathrin-mediated endocytosis pathway are key to modulating intracellular signaling and play important roles in disease. Chromosomal rearrangements that result in fusion transcripts between the Mixed-Lineage Leukemia gene (MLL) and CALM (Clathrin assembly lymphoid myeloid leukemia gene) are believed to play a role in leukemogenesis. Similarly, disruptions in Rab7 and Rab9, as well as HIP1 (Huntingtin-interacting protein), genes that are believed to be involved in endocytosis, are potentially responsible for ailments resulting in lipid storage, and neuronal diseases, respectively. For these reasons, siRNA directed

against clathrin and other genes involved in the clathrin-mediated endocytotic pathway are potentially important research and therapeutic tools.

siRNAs directed against genes involved in the clathrin-mediated endocytosis pathways were selected using Formula VIII. The targeted genes were clathrin heavy chain (CHC, accession # NM_004859), clathrin light chain A (CLCa, NM_001833), clathrin light chain B (CLCb, NM_001834), CALM (U45976), β 2 subunit of AP-2 (β 2, NM_001282), Eps15 (NM_001981), Eps15R (NM_021235), dynamin II (DYNII, NM_004945), Rab5a (BC001267), Rab5b (NM_002868), Rab5c (AF141304), and EEA.1 (XM_018197).

For each gene, four siRNAs duplexes with the highest scores were selected and a BLAST search was conducted for each of them using the Human EST database. In order to minimize the potential for off-target silencing effects, only those sequences with more than three mismatches against un-related sequences were used. All duplexes were synthesized at Dharmacon, Inc. as 21-mers with 3'-UU overhangs using a modified method of 2'-ACE chemistry Scaringe, *Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis*, Methods Enzymol 2000, 317:3 and the antisense strand was chemically phosphorylated to insure maximized activity.

HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, antibiotics and glutamine. siRNA duplexes were resuspended in 1X siRNA Universal buffer (Dharmacon, Inc.) to 20 μ M prior to transfection. HeLa cells in 12-well plates were transfected twice with 4 μ l of 20 μ M siRNA duplex in 3 μ l Lipofectamine 2000 reagent (Invitrogen, Carlsbad, California, USA) at 24-hour intervals. For the transfections in which 2 or 3 siRNA duplexes were included, the amount of each duplex was decreased, so that the total amount was the same as in transfections with single siRNAs. Cells were plated into normal culture medium 12 hours prior to experiments, and protein levels were measured 2 or 4 days after the first transfection.

Equal amounts of lysates were resolved by electrophoresis, blotted, and stained with the antibody specific to targeted protein, as well as antibodies specific to

unrelated proteins, PP1 phosphatase and Tsg101 (not shown). The cells were lysed in Triton X-100/glycerol solubilization buffer as described previously. Tebar, Bohlander, & Sorkin, *Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-mediated Traffic*, Mol. Biol. Cell Aug 1999, 10:2687. Cell lysates were electrophoresed; transferred to nitrocellulose membranes, and Western blotting was performed with several antibodies followed by detection using enhanced chemiluminescence system (Pierce, Inc). Several x-ray films were analyzed to determine the linear range of the chemiluminescence signals, and the quantifications were performed using densitometry and AlphaImager v5.5 software (Alpha Innotech Corporation). In experiments with Eps15R-targeted siRNAs, cell lysates were subjected to immunoprecipitation with Ab860, and Eps15R was detected in immunoprecipitates by Western blotting as described above.

The antibodies to assess the levels of each protein by Western blot were obtained from the following sources: monoclonal antibody to clathrin heavy chain (TD.1) was obtained from American Type Culture Collection (Rockville, MD, USA); polyclonal antibody to dynamin II was obtained from Affinity Bioreagents, Inc. (Golden, CO, USA); monoclonal antibodies to EEA.1 and Rab5a were purchased from BD Transduction Laboratories (Los Angeles, CA, USA); the monoclonal antibody to Tsg101 was purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA, USA); the monoclonal antibody to GFP was from ZYMED Laboratories Inc. (South San Francisco, CA, USA); the rabbit polyclonal antibodies Ab32 specific to α -adaptins and Ab20 to CALM were described previously Sorkin, *et al*, *Stoichiometric Interaction of the Epidermal Growth Factor Receptor with the Clathrin-associated Protein Complex AP-2*, J. Biol. Chem. Jan 1995, 270:619, the polyclonal antibodies to clathrin light chains A and B were kindly provided by Dr. F. Brodsky (UCSF); monoclonal antibodies to PP1 (BD Transduction Laboratories) and α -Actinin (Chemicon) were kindly provided by Dr. M. Dell'Acqua (University of Colorado); Eps15 Ab577 and Eps15R Ab860 were kindly provided by Dr. P.P. Di Fiore (European Cancer Institute).

Figure 11 demonstrates the *in vivo* functionality of 48 individual siRNAs, selected using Formula VIII (most of them will meet the criteria incorporated by Formula IX as well) targeting 12 genes. Various cell lines were transfected with siRNA duplexes (*Dup1-4*) or pools of siRNA duplexes (Pool), and the cells were
5 lysed 3 days after transfection with the exception of CALM (2 days) and β 2 (4 days).

Note a β 1-adaptin band (part of AP-1 Golgi adaptor complex) that runs slightly slower than β 2 adaptin. CALM has two splice variants, 66 and 72 kD. The full-length Eps15R (a doublet of ~130 kD) and several truncated spliced forms of ~
10 100 kD and ~70 kD were detected in Eps15R immunoprecipitates (shown by arrows). The cells were lysed 3 days after transfection. Equal amounts of lysates were resolved by electrophoresis and blotted with the antibody specific to a targeted protein (GFP antibody for YFP fusion proteins) and the antibody specific to unrelated proteins PP1 phosphatase or α -actinin, and TSG101. The amount of protein in each
15 specific band was normalized to the amount of non-specific proteins in each lane of the gel. Nearly all of them appear to be functional, which establishes that Formula VIII and IX can be used to predict siRNAs' functionality in general in a genome wide manner.

20 To generate the fusion of yellow fluorescent protein (YFP) with Rab5b or Rab5c (YFP-Rab5b or YFP-Rab5c), a DNA fragment encoding the full-length human Rab5b or Rab5c was obtained by PCR using *Pfu* polymerase (Stratagene) with a *SacI* restriction site introduced into the 5' end and a *KpnI* site into the 3' end and cloned into pEYFP-C1 vector (CLONTECH, Palo Alto, CA, USA). GFP-CALM and YFP-
25 Rab5a were described previously Tebar, Bohlander, & Sorkin, *Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-mediated Traffic*, Mol. Biol. Cell Aug 1999, 10:2687.

30 **Example III. Validation of the Algorithm Using Eg5, GADPH, ATE1, MEK2, MEK1, QB, LaminA/C, c-myc, human cyclophilin, and mouse cyclophilin.**

A number of genes have been identified as playing potentially important roles in disease etiology. Expression profiles of normal and diseased kidneys has implicated Edg5 in immunoglobulin A neuropathy, a common renal glomerular disease. Myc1, MEK1/2 and other related kinases have been associated with one or more cancers, while lamins have been implicated in muscular dystrophy and other diseases. For these reasons, siRNA directed against the genes encoding these classes of molecules would be important research and therapeutic tools.

Figure 12 illustrates four siRNAs targeting 10 different genes (**Table V** for sequence and accession number information) that were selected according to the Formula VIII and assayed as individuals and pools in HEK293 cells. The level of siRNA induced silencing was measured using the B-DNA assay. These studies demonstrated that thirty-six out of the forty individual SMART-selected siRNA tested are functional (90%) and all 10 pools are fully functional.

Example V. Validation of the Algorithm Using Bcl2

Bcl-2 is a ~25kD, 205-239 amino acid, anti-apoptotic protein that contains considerable homology with other members of the BCL family including BCLX, MCL1, BAX, BAD, and BIK. The protein exists in at least two forms (Bcl2a, which has a hydrophobic tail for membrane anchorage, and Bcl2b, which lacks the hydrophobic tail) and is predominantly localized to the mitochondrial membrane. While Bcl2 expression is widely distributed, particular interest has focused on the expression of this molecule in B and T cells. Bcl2 expression is down-regulated in normal germinal center B cells yet in a high percentage of follicular lymphomas, Bcl2 expression has been observed to be elevated. Cytological studies have identified a common translocation ((14;18)(q32;q32)) amongst a high percentage (>70%) of these lymphomas. This genetic lesion places the Bcl2 gene in juxtaposition to immunoglobulin heavy chain gene (IgH) encoding sequences and is believed to enforce inappropriate levels of gene expression, and resistance to programmed cell death in the follicle center B cells. In other cases, hypomethylation of the Bcl2 promoter leads to enhanced expression and again, inhibition of apoptosis. In addition to cancer, dysregulated expression of Bcl-2 has been correlated with multiple sclerosis and various neurological diseases.

The correlation between Bcl-2 translocation and cancer makes this gene an attractive target for RNAi. Identification of siRNA directed against the bcl2 transcript (or Bcl2-IgH fusions) would further our understanding Bcl2 gene function and possibly provide a future therapeutic agent to battle diseases that result from altered expression or function of this gene.

In Silico Identification of Functional siRNA

To identify functional and hyperfunctional siRNA against the Bcl2 gene, the sequence for Bcl-2 was downloaded from the NCBI Unigene database and analyzed using the Formula VIII algorithm. As a result of these procedures, both the sequence and SMARTscores™ of the Bcl2 siRNA were obtained and ranked according to their functionality. Subsequently, these sequences were BLAST'ed (database) to insure that the selected sequences were specific and contained minimal overlap with unrelated genes. The SMARTscores™ for the top 10 Bcl-2 siRNA are identified in Figure 13.

In Vivo Testing of Bcl-2 SiRNA

Bcl-2 siRNAs having the top ten SMARTscores™ were selected and tested in a functional assay to determine silencing efficiency. To accomplish this, each of the ten duplexes were synthesized using 2'-O-ACE chemistry and transfected at 100nM concentrations into cells. Twenty-four hours later assays were performed on cell extracts to assess the degree of target silencing. Controls used in these experiments included mock transfected cells, and cells that were transfected with a non-specific siRNA duplex.

The results of these experiments are presented below (and in Figure 14) and show that all ten of the selected siRNA induce 80% or better silencing of the Bcl2 message at 100nM concentrations. These data verify that the algorithm successfully identified functional Bcl2 siRNA and provide a set of functional agents that can be used in experimental and therapeutic environments.

5	siRNA 1	GGGAGAUAGUGAUGAAGUA	SEQ. ID NO. 301
	siRNA 2	GAAGUACAUCCAUUAUAAG	SEQ. ID NO. 302
	siRNA 3	GUACGACAACCGGGAGUA	SEQ. ID NO. 303
	siRNA 4	AGAUAGUGAUGAAGUACAU	SEQ. ID NO. 304
	siRNA 5	UGAAGACUCUCGUCAGUUU	SEQ. ID NO. 305
	siRNA 6	GCAUGCGGCCUCUGUUUGA	SEQ. ID NO. 306
	siRNA 7	UGCGGCCUCUGUUUGAUUU	SEQ. ID NO. 307
	siRNA 8	GAGAUAGUGAUGAAGUACA	SEQ. ID NO. 308
	siRNA 9	GGAGAUAGUGAUGAAGUAC	SEQ. ID NO. 309
	siRNA 10	GAAGACUCUGCUCAGUUUG	SEQ. ID NO. 310
10	Bcl2 siRNA: Sense Strand, 5'→3'		

Example VI. Sequences Selected by the Algorithm

- Sequences of the siRNAs selected using Formulas (Algorithms) VIII and IX with their corresponding ranking, which have been evaluated for the silencing activity *in vivo* in the present study (Formula VIII and IX, respectively).

TABLE V

Gene Name	Accession Number	SEQ. ID NO.	FTilSeqTence	Formula VIII	Formula IX
CLTC	NM_004859	SEQ. ID NO. 0301	GAAAGAATCTGTAGAGAAA	76	94.2
CLTC	NM_004859	SEQ. ID NO. 0302	GCAATGAGCTGTTTGAAGA	65	39.9
CLTC	NM_004859	SEQ. ID NO. 0303	TGACAAAGGTGGATAAATT	57	38.2
CLTC	NM_004859	SEQ. ID NO. 0304	GGAAATGGATCTCTTTGAA	54	49.4
CLTA	NM_001833	SEQ. ID NO. 0305	GGAAAGTAATGGTCCAACA	22	55.5
CLTA	NM_001833	SEQ. ID NO. 0306	AGACAGTTATGCAGCTATT	4	22.9
CLTA	NM_001833	SEQ. ID NO. 0307	CCAATTCTCGGAAGCAAGA	1	17
CLTA	NM_001833	SEQ. ID NO. 0308	GAAAGTAATGGTCCAACAG	-1	-13
CLTB	NM_001834	SEQ. ID NO. 0309	GCGCCAGAGTGAACAAGTA	17	57.5
CLTB	NM_001834	SEQ. ID NO. 0310	GAAGGTGGCCCAGCTATGT	15	-8.6
CLTB	NM_001834	SEQ. ID NO. 0311	GGAACCAGCGCCAGAGTGA	13	40.5
CLTB	NM_001834	SEQ. ID NO. 0312	GAGCGAGATTGCAGGCATA	20	61.7
CALM	U45976	SEQ. ID NO. 0313	GTTAGTATCTGATGACTTG	36	-34.6
CALM	U45976	SEQ. ID NO. 0314	GAAATGGAACCACTAAGAA	33	46.1
CALM	U45976	SEQ. ID NO. 0315	GGAAATGGAACCACTAAGA	30	61.2
CALM	U45976	SEQ. ID NO. 0316	CAACTACACTTTCCAATGC	28	6.8
EPS15	NM_001981	SEQ. ID NO. 0317	CCACCAAGATTTCATGATA	48	25.2

EPS15	NM_001981	SEQ. ID NO. 0318	GATCGGAAGCTCCAACAAGA	43	49.3
EPS15	NM_001981	SEQ. ID NO. 0319	AAACGGAGCTACAGATTAT	39	11.5
EPS15	NM_001981	SEQ. ID NO. 0320	CCACACAGCATTCTTGTA	33	-23.6
EPS15R	NM_021235	SEQ. ID NO. 0321	GAAGTTACCTTGAGCAATC	48	33
EPS15R	NM_021235	SEQ. ID NO. 0322	GGACTTGGCCGATCCAGAA	27	33
EPS15R	NM_021235	SEQ. ID NO. 0323	GCACTTGGATCGAGATGAG	20	1.3
EPS15R	NM_021235	SEQ. ID NO. 0324	CAAAGACCAATTTCGCGTTA	17	27.7
DNM2	NM_004945	SEQ. ID NO. 0325	CCGAATCAATCGCATCTTC	6	-29.6
DNM2	NM_004945	SEQ. ID NO. 0326	GACATGATCCTGCAGTTCA	5	-14
DNM2	NM_004945	SEQ. ID NO. 0327	GAGCGAATCGTCACCACTT	5	24
DNM2	NM_004945	SEQ. ID NO. 0328	CCTCCGAGCTGGCGTCTAC	-4	-63.6
ARF6	AF93885	SEQ. ID NO. 0329	TCACATGGTTAACCTCTAA	27	-21.1
ARF6	AF93885	SEQ. ID NO. 0330	GATGAGGGACGCCATAATC	7	-38.4
ARF6	AF93885	SEQ. ID NO. 0331	CCTCTAACTACAAATCTTA	4	16.9
ARF6	AF93885	SEQ. ID NO. 0332	GGAAGGTGCTATCCAAAAT	4	11.5
RAB5A	BC001267	SEQ. ID NO. 0333	GCAAGCAAGTCCTAACATT	40	25.1
RAB5A	BC001267	SEQ. ID NO. 0334	GGAAGAGGAGTAGACCTTA	17	50.1
RAB5A	BC001267	SEQ. ID NO. 0335	AGGAATCAGTGTTGTAGTA	16	11.5
RAB5A	BC001267	SEQ. ID NO. 0336	GAAGAGGAGTAGACCTTAC	12	7
RAB5B	NM_002868	SEQ. ID NO. 0337	GAAAGTCAAGCCTGGTATT	14	18.1
RAB5B	NM_002868	SEQ. ID NO. 0338	AAAGTCAAGCCTGGTATTA	6	-17.8
RAB5B	NM_002868	SEQ. ID NO. 0339	GCTATGAACGTGAATGATC	3	-21.1
RAB5B	NM_002868	SEQ. ID NO. 0340	CAAGCCTGGTATTACGTTT	-7	-37.5
RAB5C	AF141304	SEQ. ID NO. 0341	GGAACAAGATCTGTCAATT	38	51.9
RAB5C	AF141304	SEQ. ID NO. 0342	GCAATGAACGTGAACGAAA	29	43.7
RAB5C	AF141304	SEQ. ID NO. 0343	CAATGAACGTGAACGAAAT	18	43.3
RAB5C	AF141304	SEQ. ID NO. 0344	GGACAGGAGCGGTATCACA	6	18.2
EEA1	XM_018197	SEQ. ID NO. 0345	AGACAGAGCTTGAGAATAA	67	64.1
EEA1	XM_018197	SEQ. ID NO. 0346	GAGAAGATCTTTATGCAAA	60	48.7
EEA1	XM_018197	SEQ. ID NO. 0347	GAAGAGAAATCAGCAGATA	58	45.7
EEA1	XM_018197	SEQ. ID NO. 0348	GCAAGTAACTCAACTAACA	56	72.3
AP2B1	NM_001282	SEQ. ID NO. 0349	GAGCTAATCTGCCACATTG	49	-12.4
AP2B1	NM_001282	SEQ. ID NO. 0350	GCAGATGAGTTACTAGAAA	44	48.9
AP2B1	NM_001282	SEQ. ID NO. 0351	CAACTTAATTGTCCAGAAA	41	28.2
AP2B1	NM_001282	SEQ. ID NO. 0352	CAACACAGGATTCTGATAA	33	-5.8
PLK	NM_005030	SEQ. ID NO. 0353	AGATTGTGCCTAAGTCTCT	-35	-3.4
PLK	NM_005030	SEQ. ID NO. 0354	ATGAAGATCTGGAGGTGAA	0	-4.3
PLK	NM_005030	SEQ. ID NO. 0355	TTTGAGACTTCTTGCCTAA	-5	-27.7
PLK	NM_005030	SEQ. ID NO. 0356	AGATCACCCCTCCTTAAATA	15	72.3
GAPDH	NM_002046	SEQ. ID NO. 0357	CAACGGATTGCGTATT	27	-2.8

GAPDH	NM_002046	SEQ. ID NO. 0358	GAAATCCCATCACCATCTT	24	3.9
GAPDH	NM_002046	SEQ. ID NO. 0359	GACCTCAACTACATGGTTT	22	-22.9
GAPDH	NM_002046	SEQ. ID NO. 0360	TGGTTTACATGTTCCAATA	9	9.8
c-Myc		SEQ. ID NO. 0361	GAAGAAATCGATGTTGTTT	31	-11.7
c-Myc		SEQ. ID NO. 0362	ACACAAACTTGAACAGCTA	22	51.3
c-Myc		SEQ. ID NO. 0363	GGAAGAAATCGATGTTGTT	18	26
c-Myc		SEQ. ID NO. 0364	GAAACGACGAGAACAGTTG	18	-8.9
MAP2K1	NM_002755	SEQ. ID NO. 0365	GCACATGGATGGAGGTTCT	26	16
MAP2K1	NM_002755	SEQ. ID NO. 0366	GCAGAGAGAGCAGATTTGA	16	0.4
MAP2K1	NM_002755	SEQ. ID NO. 0367	GAGGTTCTCTGGATCAAGT	14	15.5
MAP2K1	NM_002755	SEQ. ID NO. 0368	GAGCAGATTTGAAGCAACT	14	18.5
MAP2K2	NM_030662	SEQ. ID NO. 0369	CAAAGACGATGACTTCGAA	37	26.4
MAP2K2	NM_030662	SEQ. ID NO. 0370	GATCAGCATTTCATGGAA	24	-0.7
MAP2K2	NM_030662	SEQ. ID NO. 0371	TCCAGGAGTTTGTCAATAA	17	-4.5
MAP2K2	NM_030662	SEQ. ID NO. 0372	GGAAGCTGATCCACCTTGA	16	59.2
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0373	GCAGAAATCTAAGGATATA	53	35.8
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0374	CAACAAGGATGAAGTCTAT	50	18.3
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0375	CAGCAGAAATCTAAGGATA	41	32.7
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0376	CTAGATGGCTTTCTCAGTA	39	3.9
CyclophilinA_	NM_021130	SEQ. ID NO. 0377	AGACAAGGTCCCAAAGACA	-16	58.1
CyclophilinA_	NM_021130	SEQ. ID NO. 0378	GGAATGGCAAGACCAGCAA	-6	36
CyclophilinA_	NM_021130	SEQ. ID NO. 0379	AGAATTATTCCAGGGTTTA	-3	16.1
CyclophilinA_	NM_021130	SEQ. ID NO. 0380	GCAGACAAGGTCCCAAAGA	8	8.9
LAMIN A/C	NM_170707	SEQ. ID NO. 0381	AGAAGCAGCTTCAGGATGA	31	38.8
LAMIN A/C	NM_170707	SEQ. ID NO. 0382	GAGCTTGACTTCCAGAAGA	33	22.4
LAMIN A/C	NM_170707	SEQ. ID NO. 0383	CCACCGAAGTTCACCCTAA	21	27.5
LAMIN A/C	NM_170707	SEQ. ID NO. 0384	GAGAAGAGCTCCTCCATCA	55	30.1
CyclophilinB	M60857	SEQ. ID NO. 0385	GAAAGAGCATCTACGGTGA	41	83.9
CyclophilinB	M60857	SEQ. ID NO. 0386	GAAAGGATTTGGCTACAAA	53	59.1
CyclophilinB	M60857	SEQ. ID NO. 0387	ACAGCAAATTCCATCGTGT	-20	28.8
CyclophilinB	M60857	SEQ. ID NO. 0388	GGAAAGACTGTTCCAAAAA	2	27
DBI1	NM_020548	SEQ. ID NO. 0389	CAACACGCCTCATCCTCTA	27	-7.6
DBI2	NM_020548	SEQ. ID NO. 0390	CATGAAAGCTTACATCAAC	25	-30.8
DBI3	NM_020548	SEQ. ID NO. 0391	AAGATGCCATGAAAGCTTA	17	22
DBI4	NM_020548	SEQ. ID NO. 0392	GCACATACCGCCTGAGTCT	15	3.9
rLUC1		SEQ. ID NO. 0393	GATCAAATCTGAAGAAGGA	57	49.2
rLUC2		SEQ. ID NO. 0394	GCCAAGAAGTTTCCTAATA	50	13.7
rLUC3		SEQ. ID NO. 0395	CAGCATATCTTGAACCATT	41	-2.2
rLUC4		SEQ. ID NO. 0396	GAACAAAGGAAACGGATGA	39	29.2
SeAP1	NM_031313	SEQ. ID NO. 0397	CGGAAACGGTCCAGGCTAT	6	26.9

SeAP2	NM_031313	SEQ. ID NO. 0398	GCTTCGAGCAGACATGATA	4	-11.2
SeAP3	NM_031313	SEQ. ID NO. 0399	CCTACACGGTCCTCCTATA	4	4.9
SeAP4	NM_031313	SEQ. ID NO. 0400	GCCAAGAACCTCATCATCT	1	-9.9
fLUC1		SEQ. ID NO. 0401	GATATGGGCTGAATACAAA	54	40.4
fLUC2		SEQ. ID NO. 0402	GCACTCTGATTGACAAATA	47	54.7
fLUC3		SEQ. ID NO. 0403	TGAAGTCTCTGATTAAGTA	46	34.5
fLUC4		SEQ. ID NO. 0404	TCAGAGAGATCCTCATAAA	40	11.4
mCyclo_1	NM_008907	SEQ. ID NO. 0405	GCAAGAAGATCACCATTTTC	52	46.4
mCyclo_2	NM_008907	SEQ. ID NO. 0406	GAGAGAAATTTGAGGATGA	36	70.7
mCyclo_3	NM_008907	SEQ. ID NO. 0407	GAAAGGATTTGGCTATAAG	35	-1.5
mCyclo_4	NM_008907	SEQ. ID NO. 0408	GAAAGAAGGCATGAACATT	27	10.3
BCL2_1	NM_000633	SEQ. ID NO. 0409	GGGAGATAGTGATGAAGTA	21	72
BCL2_2	NM_000633	SEQ. ID NO. 0410	GAAGTACATCCATTATAAG	1	3.3
BCL2_3	NM_000633	SEQ. ID NO. 0411	GTACGACAACCGGGAGATA	1	35.9
BCL2_4	NM_000633	SEQ. ID NO. 0412	AGATAGTGATGAAGTACAT	-12	22.1
BCL2_5	NM_000633	SEQ. ID NO. 0413	TGAAGACTCTGCTCAGTTT	36	19.1
BCL2_6	NM_000633	SEQ. ID NO. 0414	GCATGCGGCCTCTGTTTGA	5	-9.7
QB1	NM_003365.1	SEQ. ID NO. 0415	GCACACAGCUUACUACAUC	52	-4.8
QB2	NM_003365.1	SEQ. ID NO. 0416	GAAAUGCCCUGGUUAUCUCA	49	22.1
QB3	NM_003365.1	SEQ. ID NO. 0417	GAAGGAACGUGAUGUGAUC	34	22.9
QB4	NM_003365.1	SEQ. ID NO. 0418	GCACUACUCCUGUGUGUGA	28	20.4
ATE1-1	NM_007041	SEQ. ID NO. 0419	GAACCCAGCUGGAGAACUU	45	15.5
ATE1-2	NM_007041	SEQ. ID NO. 0420	GAUAUACAGUGUGAUCUUA	40	12.2
ATE1-3	NM_007041	SEQ. ID NO. 0421	GUACUACGAUCCUGAUUUAU	37	32.9
ATE1-4	NM_007041	SEQ. ID NO. 0422	GUGCCGACCUUUACAAUUU	35	18.2
EGFR-1	NM_005228	SEQ. ID NO. 0423	GAAGGAAACTGAATTCAAA	68	79.4
EGFR-1	NM_005228	SEQ. ID NO. 0424	GGAAATATGTACTACGAAA	49	49.5
EGFR-1	NM_005228	SEQ. ID NO. 0425	CCACAAAGCAGTGAATTTA	41	7.6
EGFR-1	NM_005228	SEQ. ID NO. 0426	GTAACAAGCTCACGCAGTT	40	25.9

Example VII. Genome-Wide Application of the Algorithm

The examples described above demonstrate that the algorithm(s) can successfully identify functional siRNA and that these duplexes can be used to induce the desirable phenotype of transcriptional knockdown or knockout. Each gene or family of genes in each organism plays an important role in maintaining physiological homeostasis and the algorithm can be used to develop functional, highly functional, or hyperfunctional siRNA to each gene. To accomplish this for the human genome, the entire online ncbi refseq database was accessed through Entrez (efetch). The database

was processed through Formula VIII. For each gene the top 80 –100 scores for siRNAs were obtained and BLAST'ed to insure that the selected sequences are specific in targeting the gene of choice. These sequences are provided on the enclosed CDs in electronic form.

5

With respect to the disks, there are four tables on each disk copy in text format: Tables XII –XV. Table XII, which is located in a file entitled Table_12.txt, provides a list of the 80-100 sequences for each target, identified by Formula VIII as having the highest relative SMARTscoresTM for the target analyzed. Table XIII, which is located in a file entitled Table_13.txt, provides the SMARTscoresTM, and for each gene, a pool pick of up to four sequences is denoted. (The denotation of "1" in Table XIII means that it is a pool pick.) These pool pick sequences represent the most functional siRNAs for the corresponding target. Any 1, 2, 3, or 4 of the pool pick sequences could be used for gene silencing. Further, sequences that are not denoted as pool pick sequences, but that are included on the compact disks may also be used for gene silencing either alone or in combination with other sequences. However, their individual relative functionality would be less than that of a pool pick sequence. Table XIV, which is located in a file entitled Table_14.txt, provides an identification of genes by accession number, and Table XV, which is located in a file entitled Table_15.txt, provides a short name for the genes identified on the disk. The information contained on the disks is part of this patent application and are incorporated into the specification by reference. One may use these tables in order to identify functional siRNAs for the gene provided therein, by simply looking for the gene of interest and an siRNA that is listed as functional. Preferably, one would select one or more of the siRNA that most optimized for the target of interest and is denoted as a pool pick.

Table XII: siRNA selected by Formula VIII

See data submitted herewith on a CD-ROM in accordance with PCT
Administrative Instructions Section 801(a)

Table XIII: SMARTscoresTM

See data submitted herewith on a CD-ROM in accordance with PCT
Administrative Instructions Section 801(a)

Table XIV: Identification of Targets

See data submitted herewith on a CD-ROM in accordance with PCT
Administrative Instructions Section 801(a)

5

Table XV: Description of Targets

See data submitted herewith on a CD-ROM in accordance with PCT
Administrative Instructions Section 801(a)

10 Many of the genes to which the described siRNA are directed play critical
roles in disease etiology. For this reason, the siRNA listed in the accompanying
compact disk may potentially act as therapeutic agents. A number of prophetic
examples follow and should be understood in view of the siRNA that are identified on
the accompanying CD. To isolate these siRNA, the appropriate message sequence for
15 each gene is analyzed using one of the before mentioned formulas (preferably formula
VIII) to identify potential siRNA targets. Subsequently these targets are BLAST'ed to
eliminate homology with potentially off-targets.

The list of potential disease targets is extensive. For instance, over-expression
20 of Bcl10 has been implicated in the development of MALT lymphoma (mucosa
associated lymphoid tissue lymphoma) and thus, functional, highly functional, or
hyperfunctional siRNA directed against that gene (*e.g.* SEQ. ID NO. 0427:
GGAAACCUCUCAUUGCUAA; SEQ. ID NO. 0428:
GAAAGAACCUUGCCGAUCA; SEQ. ID NO. 0429:
25 GGAAAUACAUCAGAGCUUA, or SEQ. ID NO. 0430:
GAAAGUAUGUGUCUUAAGU) may contribute to treatment of this disorder.

In another example, studies have shown that molecules that inhibit
glutamine:fructose-6-phosphate aminotransferase (GFA) may act to limit the
30 symptoms suffered by Type II diabetics. Thus, functional, highly functional, or
hyperfunctional siRNA directed against GFA (also known as GFPT1: siRNA = SEQ.
ID NO. 0433 UGAAACGGCUGCCUGAUUU; SEQ. ID NO. 0434
GAAGUUACCUCUUACAUUU; SEQ. ID NO. 0435

GUACGAAACUGUAUGAUUA; SEQ. ID NO. 0436

GGACGAGGCUAUCAUUAUG) may contribute to treatment of this disorder.

In another example, the von Hippel-Lindau (VHL) tumor suppressor has been
5 observed to be inactivated at a high frequency in sporadic clear cell renal cell
carcinoma (RCC) and RCCs associated with VHL disease. The VHL tumor
suppressor targets hypoxia-inducible factor-1 alpha (HIF-1 alpha), a transcription
factor that can induce vascular endothelial growth factor (VEGF) expression, for
ubiquitination and degradation. Inactivation of VHL can lead to increased levels of
10 HIF-1 alpha, and subsequent VEGF over expression. Such over expression of VEGF
has been used to explain the increased (and possibly necessary) vascularity observed
in RCC. Thus, functional, highly functional, or hyperfunctional siRNAs directed
against either HIF-1 alpha (SEQ. ID NO. 0437 GAAGGAACCUGAUGCUUUA;
SEQ. ID NO. 0438 GCAUAUAUCUAGAAGGUAU; SEQ. ID NO. 0439
15 GAACAAAUACAUGGGAUUA; SEQ. ID NO. 0440
GGACACAGAUUUAGACUUG) or VEGF (SEQ. ID NO. 0441
GAACGUACUUGCAGAUGUG; SEQ. ID NO. 0442
GAGAAAGCAUUUGUUUGUA; SEQ. ID NO. 0443
GGAGAAAGCAUUUGUUUGU; SEQ. ID NO. 0444
20 CGAGGCAGCUUGAGUUA) may be useful in the treatment of renal cell
carcinoma.

In another example, gene expression of platelet derived growth factor A and B
(PDGF-A and PDGF-B) has been observed to be increased 22- and 6-fold,
25 respectively, in renal tissues taken from patients with diabetic nephropathy as
compared with controls. These findings suggest that over expression of PDGF A and
B may play a role in the development of the progressive fibrosis that characterizes
human diabetic kidney disease. Thus, functional, highly functional, or hyperfunctional
siRNAs directed against either PDGF A
30 (SEQ. ID NO. 0445: GGUAAGAUUUGUGCUUUA;
SEQ. ID NO. 0446: CCGCAAAUAUGCAGAAUUA;
SEQ. ID NO. 0447: GGAUGUACAUGGCGUGUUA;
SEQ. ID NO. 0448: GGUGAAGUUUGUAUGUUUA) or

PDGF B

(SEQ. ID NO. 0449: CCGAGGAGCUUUAUGAGAU;

SEQ. ID NO. 0450: GCUCCGCGCUUCCGAUUU;

SEQ. ID NO. 0451 GAGCAGGAAUGGUGAGAUG;

5 SEQ. ID NO. 0452: GAACUUGGGAUAAGAGUGU;

SEQ. ID NO. 0453 CCGAGGAGCUUUAUGAGAU;

SEQ. ID NO. 0454 UUUAUGAGAUGCUGAGUGA) may be useful in the treatment of this form of kidney disorder.

10 In another example, a strong correlation exists between the over-expression of glucose transporters (*e.g.* GLUT12) and cancer cells. It is predicted that cells undergoing uncontrolled cell growth up-regulate GLUT molecules so that they can cope with the heightened energy needs associated with increased rates of proliferation and metastasis. Thus, siRNA-based therapies that target the molecules such as

15 GLUT1 (also known as SLC2A1: siRNA=

SEQ. ID NO.: 0455 GCAAUGAUGUCCAGAAGAA;

SEQ. ID NO.: 0456 GAAGAAUAUUCAGGACUUA;

SEQ. ID NO.: 0457 GAAGAGAGUCGGCAGAUGA;

SEQ. ID NO.: 0458 CCAAGAGUGUGCUAAAGAA)

20

GLUT12 (also known as SLC12: siRNA =

SEQ. ID NO. 0459: GAGACACUCUGAAAUGAUA;

SEQ. ID NO. 0460: GAAAUGAUGUGGAUAAGAG;

SEQ. ID NO. 0461: GAUCAAUCCUCCCUGAAA;

25 SEQ. ID NO. 0462: UGAAUGAGCUGAUGAUUGU) and other related transporters, may be of value in treating a multitude of malignancies.

The siRNA sequences listed above are presented in a 5' → 3' sense strand direction. In addition, siRNA directed against the targets listed above as well as those directed against other targets and listed in the accompanying compact disk may be
30 useful as therapeutic agents.

Example VIII. Evidence for the Benefits of Pooling

Evidence for the benefits of pooling have been demonstrated using the reporter gene, luciferase. Ninety siRNA duplexes were synthesized using Dharmacon proprietary ACE® chemistry against one of the standard reporter genes: firefly luciferase. The duplexes were designed to start two base pairs apart and to cover approximately 180 base pairs of the luciferase gene (see sequences in **Table III**). Subsequently, the siRNA duplexes were co-transfected with a luciferase expression reporter plasmid into HEK293 cells using standard transfection protocols and luciferase activity was assayed at 24 and 48 hours.

Transfection of individual siRNAs showed standard distribution of inhibitory effect. Some duplexes were active, while others were not. **Figure 15** represents a typical screen of ninety siRNA duplexes (SEQ. ID NO. 0032- 0120) positioned two base pairs apart. As the figure suggests, the functionality of the siRNA duplex is determined more by a particular sequence of the oligonucleotide than by the relative oligonucleotide position within a gene or excessively sensitive part of the mRNA, which is important for traditional anti-sense technology.

When two continuous oligonucleotides were pooled together, a significant increase in gene silencing activity was observed. (See **Figure 16**) A gradual increase in efficacy and the frequency of pools functionality was observed when the number of siRNAs increased to 3 and 4. (**Figures 16, 17**). Further, the relative positioning of the oligonucleotides within a pool did not determine whether a particular pool was functional (see **Figure 18**, in which 100% of pools of oligonucleotides distanced by 2, 10 and 20 base pairs were functional).

However, relative positioning may nonetheless have an impact. An increased functionality may exist when the siRNA are positioned continuously head to toe (5' end of one directly adjacent to the 3' end of the others).

Additionally, siRNA pools that were tested performed at least as well as the best oligonucleotide in the pool, under the experimental conditions whose results are depicted in **Figure 19**. Moreover, when previously identified non-functional and marginally (semi) functional siRNA duplexes were pooled together in groups of five at a time, a significant functional cooperative action was observed. (See **Figure 20**)

In fact, pools of semi-active oligonucleotides were 5 to 25 times more functional than the most potent oligonucleotide in the pool. Therefore, pooling several siRNA duplexes together does not interfere with the functionality of the most potent siRNAs within a pool, and pooling provides an unexpected significant increase in overall
5 functionality

Example IX. Pooling Across Species

Experiments were performed on the following genes: β -galactosidase, Renilla luciferase, and Secreted alkaline phosphatase, which demonstrates the benefits of
10 pooling. (see **Figure 21**) Approximately 50% of individual siRNAs designed to silence the above-specified genes were functional, while 100% of the pools that contain the same siRNA duplexes were functional.

Example X. Highly Functional siRNA

15 Pools of five siRNAs in which each two siRNAs overlap to 10-90% resulted in 98% functional entities (>80% silencing). Pools of siRNAs distributed throughout the mRNA that were evenly spaced, covering an approximate 20 – 2000 base pair range, were also functional. When the pools of siRNA were positioned continuously head to tail relative to mRNA sequences and mimicked the natural products of Dicer
20 cleaved long double stranded RNA, 98% of the pools evidenced highly functional activity (>95% silencing).

Example XI. Human cyclophyline

Table III above lists the siRNA sequences for the human cyclophyline
25 protein. A particularly functional siRNA may be selected by applying these sequences to any of Formula I to VII above.

Alternatively, one could pool 2, 3, 4, 5 or more of these sequences to create a kit for silencing a gene. Preferably, within the kit there would be at least one
30 sequence that has a relatively high predicted functionality when any of Formulas I - VII is applied.

Example XII. Sample Pools of siRNAs and Their Application to Human Disease

The genetic basis behind human disease is well documented and siRNA may be used as both research or diagnostic tools and therapeutic agents, either individually or in pools. Genes involved in signal transduction, the immune response, apoptosis, DNA repair, cell cycle control, and a variety of other physiological functions have clinical relevance and therapeutic agents that can modulate expression of these genes may alleviate some or all of the associated symptoms. In some instances, these genes can be described as a member of a family or class of genes and siRNA (randomly, conventionally, or rationally designed) can be directed against one or multiple members of the family to induce a desired result.

To identify rationally designed siRNA to each gene, the sequence was analyzed using Formula-VIII to identify a SMARTpool containing the functional sequences. To confirm the activity of these sequences, the siRNA are introduced into a cell type of choice (e.g. HeLa cells, HEK293 cells) and the levels of the appropriate message are analyzed using one of several art proven techniques. SiRNA having heightened levels of potency can be identified by testing each of the before mentioned duplexes at increasingly limiting concentrations. Similarly, siRNA having increased levels of longevity can be identified by introducing each duplex into cells and testing functionality at 24, 48, 72, 96, 120, 144, 168, and 192 hours after transfection. Agents that induce >95% silencing at sub-nanomolar concentrations and/or induce functional levels of silencing for >96 hours are considered hyperfunctional.

The following are non-limiting examples of families of proteins to which siRNA described in this document are targeted against:

Transporters, Pumps, and Channels

Transporters, pumps, and channels represent one class of genes that are attractive targets for siRNAs. One major class of transporter molecules are the ATP-binding cassette (ABC) transporters. To date, nearly 50 human ABC-transporter genes have been characterized and have been shown to be involved in a variety of physiological functions including transport of bile salts, nucleosides, chloride ions, cholesterol, toxins, and more. Predominant among this group are MDR1 (which encodes the P-glycoprotein, NP_000918), the MDR-related proteins (MRP1-7), and

the breast cancer resistance protein (BCRP). In general, these transporters share a common structure, with each protein containing a pair of ATP-binding domains (also known as nucleotide binding folds, NBF) and two sets of transmembrane (TM) domains, each of which typically contains six membrane-spanning α -helices. The genes encoding this class of transporter are organized as either full transporters (*i.e.* containing two TM and two NBF domains) or as half transporters that assemble as either homodimers or heterodimers to create functional transporters. As a whole, members of the family are widely dispersed throughout the genome and show a high degree of amino acid sequence identity among eukaryotes.

10

ABC-transporters have been implicated in several human diseases. For instance, molecular efflux pumps of this type play a major role in the development of drug resistance exhibited by a variety of cancers and pathogenic microorganisms. In the case of human cancers, increased expression of the MDR1 gene and related pumps have been observed to generate drug resistance to a broad collection of commonly used chemotherapeutics including doxorubicin, daunorubicin, vinblastine, vincristine, colchicines. In addition to the contribution these transporters make to the development of multi-drug resistance, there are currently 13 human genetic diseases associated with defects in 14 different transporters. The most common of these conditions include cystic fibrosis, Stargardt disease, age-related macular degeneration, adrenoleukodystrophy, Tangier disease, Dubin-Johnson syndrome and progressive familial intrahepatic cholestasis. For this reason, siRNAs directed against members of this, and related, families are potentially valuable research and therapeutic tools.

15

With respect to channels, analysis of *Drosophila* mutants has enabled the initial molecular isolation and characterization of several distinct channels including (but not limited to) potassium (K^+) channels. This list includes shaker (Sh), which encodes a voltage activated K^+ channel, slowpoke (Slo), a Ca^{2+} activated K^+ channel, and ether-a-go-go (Eag). The Eag family is further divided into three subfamilies: Eag, Elk (eag-like K channels), and Erg (Eag related genes).

25

30

The Erg subfamily contains three separate family members (Erg1-3) that are distantly related to the sh family of voltage activated K^+ channels. Like sh, erg

polypeptides contain the classic six membrane spanning architecture of K^+ channels (S1-S6) but differ in that each includes a segment associated with the C-terminal cytoplasmic region that is homologous to cyclic nucleotide binding domains (cNBD). Like many isolated ion channel mutants, erg mutants are temperature-sensitive
5 paralytics, a phenotype caused by spontaneous repetitive firing (hyperactivity) in neurons and enhanced transmitter release at the neuromuscular junction.

Initial studies on the tissue distribution of all three members of the erg subfamily show two general patterns of expression. Erg1 and erg3 are broadly
10 expressed throughout the nervous system and are observed in the heart, the superior mesenteric ganglia, the celiac ganglia, the retina, and the brain. In contrast, erg2 shows a much more restricted pattern of expression and is only observed in celiac ganglia and superior mesenteric ganglia. Similarly, the kinetic properties of the three erg potassium channels are not homogeneous. Erg1 and erg2 channels are relatively
15 slow activating delayed rectifiers whereas the erg3 current activates rapidly and then exhibits a predominantly transient component that decays to a sustained plateau. The current properties of all three channels are sensitive to methanesulfonanilides, suggesting a high degree of conservation in the pore structure of all three proteins.

20 Recently, the erg family of K^+ channels has been implicated in human disease. Consistent with the observation that erg1 is expressed in the heart, single strand conformation polymorphism and DNA sequence analyses have identified HERG (human erg1) mutations in six long-QT-syndrome (LQT) families, an inherited disorder that results in sudden death from a ventricular tachyarrhythmia. Thus siRNA
25 directed against this group of molecules (e.g. KCNH1-8) will be of extreme therapeutic value.

Another group of channels that are potential targets of siRNAs are the CLCA family that mediate a Ca^{2+} -activated Cl^- conductance in a variety of
30 tissues. To date, two bovine (bCLC1; bCLCA2 (Lu-ECAM-1)), three mouse (mCLCA1; mCLCA2; mCLCA3) and four human (hCLCA1; hCLCA2; hCLCA3; hCLCA4) CLCA family members have been isolated and patch-clamp studies with transfected human embryonic kidney (HEK-293) cells have shown that bCLCA1, mCLCA1, and hCLCA1 mediate a Ca^{2+} -activated Cl^- conductance that can be

inhibited by the anion channel blocker DIDS and the reducing agent dithiothreitol (DTT).

The protein size, structure, and processing seem to be similar among different CLCA family members and has been studied in greatest detail for Lu-ECAM-1. The Lu-ECAM-1 open reading frame encodes a precursor glycoprotein of 130 kDa that is processed to a 90-kDa amino-terminal cleavage product and a group of 30- to 40-kDa glycoproteins that are glycosylation variants of a single polypeptide derived from its carboxy terminus. Both subunits are associated with the outer cell surface, but only the 90-kDa subunit is thought to be anchored to the cell membrane via four transmembrane domains.

Although the protein processing and function appear to be conserved among CLCA homologs, significant differences exist in their tissue expression patterns. For example, bovine Lu-ECAM-1 is expressed primarily in vascular endothelia, bCLCA1 is exclusively detected in the trachea, and hCLCA1 is selectively expressed in a subset of human intestinal epithelial cells. Thus the emerging picture is that of a multigene family with members that are highly tissue specific, similar to the CIC family of voltage-gated Cl^- channels. The human channel, hCLCA2, is particularly interesting from a medical and pharmacological standpoint. CLCA2 is expressed on the luminal surface of lung vascular endothelia and serves as an adhesion molecule for lung metastatic cancer cells, thus mediating vascular arrest and lung colonization. Expression of this molecule in normal mammary epithelium is consistently lost in human breast cancer and in nearly all tumorigenic breast cancer cell lines. Moreover, re-expression of hCLCA2 in human breast cancer cells abrogates tumorigenicity in nude mice, implying that hCLCA2 acts as a tumour suppressor in breast cancer. For these reasons, siRNA directed against CLCA family members and related channels may prove to be valuable in research and therapeutic venues.

Transporters Involved in Synaptic Transmission.

Synaptic transmission involves the release of a neurotransmitter into the synaptic cleft, interaction of that transmitter with a postsynaptic receptor, and subsequent removal of the transmitter from the cleft. In most synapses the signal is terminated by a rapid reaccumulation of the neurotransmitter into presynaptic

terminals. This process is catalyzed by specific neurotransmitter transporters that are often energized by the electrochemical gradient of sodium across the plasma membrane of the presynaptic cells.

5 Aminobutyric acid (GABA) is the major inhibitory neurotransmitter in the central nervous system. The inhibitory action of GABA, mediated through GABA_A/GABA_B receptors, and is regulated by GABA transporters (GATs), integral membrane proteins located perisynaptically on neurons and glia. So far four different carriers (GAT1-GAT4) have been cloned and their cellular distribution has been
10 partly worked out. Comparative sequence analysis has revealed that GABA transporters are related to several other proteins involved in neurotransmitter uptake including gamma-aminobutyric acid transporters, monoamine transporters, amino acid transporters, certain "orphan" transporters, and the recently discovered bacterial transporters. Each of these proteins has a similar 12 transmembrane helices topology
15 and relies upon the Na⁺/Cl⁻ gradient for transport function. Transport rates are dependent on substrate concentrations, with half-maximal effective concentrations for transport frequently occurring in the submicromolar to low micromolar range. In addition, transporter function is bidirectional, and non-vesicular efflux of transmitter may contribute to ambient extracellular transmitter levels.

20 Recent evidence suggests that GABA transporters, and neurotransmitter transporters in general, are not passive players in regulating neuronal signaling; rather, transporter function can be altered by a variety of initiating factors and signal transduction cascades. In general, this functional regulation occurs in two ways,
25 either by changing the rate of transmitter flux through the transporter or by changing the number of functional transporters on the plasma membrane. A recurring theme in transporter regulation is the rapid redistribution of the transporter protein between intracellular locations and the cell surface. In general, this functional modulation occurs in part through activation of second messengers such as kinases, phosphatases,
30 arachidonic acid, and pH. However, the mechanisms underlying transporter phosphorylation and transporter redistribution have yet to be fully elucidated.

GABA transporters play a pathophysiological role in a number of human diseases including temporal lobe epilepsy and are the targets of pharmacological

interventions. Studies in seizure sensitive animals show some (but not all) of the GAT transporters have altered levels of expression at times prior to and post seizure, suggesting this class of transporter may affect epileptogenesis, and that alterations following seizure may be compensatory responses to modulate seizure activity. For these reasons, siRNAs directed against members of this family of genes (including but not limited to SLC6A1-12) may prove to be valuable research and therapeutic tools.

Organic Ion Transporters.

The human body is continuously exposed to a great variety of xenobiotics, via food, drugs, occupation, and environment. Excretory organs such as kidney, liver, and intestine defend the body against the potentially harmful effects of these compounds by transforming them into less active metabolites that are subsequently secreted from the system.

Carrier-mediated transport of xenobiotics and their metabolites exist for the active secretion of organic anions and cations. Both systems are characterized by a high clearance capacity and tremendous diversity of substances accepted, properties that result from the existence of multiple transporters with overlapping substrate specificities. The class of organic anion transporters plays a critical role in the elimination of a large number of drugs (*e.g.*, antibiotics, chemotherapeutics, diuretics, nonsteroidal anti-inflammatory drugs, radiocontrast agents, cytostatics); drug metabolites (especially conjugation products with glutathione, glucuronide, glycine, sulfate, acetate); and toxicants and their metabolites (*e.g.*, mycotoxins, herbicides, plasticizers, glutathione *S*-conjugates of polyhaloalkanes, polyhaloalkenes, hydroquinones, aminophenols), many of which are specifically harmful to the kidney.

Over the past couple of years the number of identified anion transporting molecules has grown tremendously. Uptake of organic anions (OA^-) across the basolateral membrane is mediated by the classic sodium-dependent organic anion transport system, which includes α -ketoglutarate ($\alpha\text{-KG}^{2-}$)/ OA^- exchange via the organic anion transporter (OAT1) and sodium-ketoglutarate cotransport via the Na^+ /dicarboxylate cotransporter (SDCT2). The organic anion transporting polypeptide, Oatp1, and the kidney-specific OAT-K1 and OAT-K2 are seen as potential molecules that mediate facilitated OA^- efflux but could also be involved in reabsorption via an

exchange mechanism. Lastly the PEPT1 and PEPT2 mediate luminal uptake of peptide drugs, whereas CNT1 and CNT2 are involved in reabsorption of nucleosides

The organic anion-transporting polypeptide 1 (Oatp1) is a Na⁺- and ATP-independent transporter originally cloned from rat liver. The tissue distribution and transport properties of the Oatp1 gene product are complex. Oatp1 is localized to the basolateral membrane of hepatocytes, and is found on the apical membrane of S3 proximal tubules. Studies with transiently transfected cells (e.g. HeLa cells) have indicated that Oatp1 mediates transport of a variety of molecules including taurocholate, estrone-3-sulfate, aldosterone, cortisol, and others. The observed uptake of taurocholate by Oatp1 expressed in *X. laevis* oocytes is accompanied by efflux of GSH, suggesting that transport by this molecule may be glutathione dependent.

Computer modeling suggests that members of the Oatp family are highly conserved, hydrophobic, and have 12 transmembrane domains. Decreases in expression of Oatp family members have been associated with cholestatic liver diseases and human hepatoblastomas, making this family of proteins of key interest to researchers and the medical community. For these reasons, siRNAs directed against OAT family members (including but not limited to SLC21A2, 3, 6, 8, 9, 11, 12, 14, 15, and related transporters) are potentially useful as research and therapeutic tools.

Nucleoside transporters.

Nucleoside transporters play key roles in physiology and pharmacology. Uptake of exogenous nucleosides is a critical first step of nucleotide synthesis in tissues such as bone marrow and intestinal epithelium and certain parasitic organisms that lack *de novo* pathways for purine biosynthesis. Nucleoside transporters also control the extracellular concentration of adenosine in the vicinity of its cell surface receptors and regulate processes such as neurotransmission and cardiovascular activity. Adenosine itself is used clinically to treat cardiac arrhythmias, and nucleoside transport inhibitors such as dipyridamole, dilazep, and draflazine function as coronary vasodilators.

In mammals, plasma membrane transport of nucleosides is brought about by members of the concentrative, Na⁺-dependent (CNT) and equilibrative, Na⁺-

independent (ENT) nucleoside transporter families. CNTs are expressed in a tissue-specific fashion; ENTs are present in most, possibly all, cell types and are responsible for the movement of hydrophilic nucleosides and nucleoside analogs down their concentration gradients. In addition, structure/function studies of ENT family members have predicted these molecules to contain eleven transmembrane helical segments with an amino terminus that is intracellular and a carboxyl terminus that is extracellular. The proteins have a large glycosylated loop between TMs 1 and 2 and a large cytoplasmic loop between TMs 6 and 7. Recent investigations have implicated the TM 3-6 region as playing a central role in solute recognition. The medical importance of the ENT family of proteins is broad. In humans adenosine exerts a range of cardioprotective effects and inhibitors of ENTs are seen as being valuable in alleviating a variety of cardio/cardiovascular ailments. In addition, responses to nucleoside analog drugs has been observed to vary considerably amongst *e.g.* cancer patients. While some forms of drug resistance have been shown to be tied to the up-regulation of ABC-transporters (*e.g.* MDR1), resistance may also be the result of reduced drug uptake (*i.e.* reduced ENT expression). Thus, a clearer understanding of ENT transporters may aid in optimizing drug treatments for patients suffering a wide range of malignancies. For these reasons, siRNAs directed against this class of molecules (including SLC28A1-3, SLC29A1-4, and related molecules) may be useful as therapeutic and research tools.

Sulfate Transporters.

All cells require inorganic sulfate for normal function. Sulfate is the fourth most abundant anion in human plasma and is the major source of sulfur in many organisms. Sulfation of extracellular matrix proteins is critical for maintaining normal cartilage metabolism and sulfate is an important constituent of myelin membranes found in the brain

Because sulfate is a hydrophilic anion that cannot passively cross the lipid bilayer of cell membranes, all cells require a mechanism for sulfate influx and efflux to ensure an optimal supply. To date, a variety of sulfate transporters have been identified in tissues from many origins. These include the renal sulfate transporters (NaSi-1 and Sat-1), the ubiquitously expressed diastrophic dysplasia sulfate transporter (DTDST), the intestinal sulfate transporter (DRA), and the erythrocyte

anion exchanger (AE1). Most, if not all, of these molecules contain the classic 12 transmembrane spanning domain architecture commonly found amongst members of the anion transporter superfamily.

5 Recently three different sulfate transporters have been associated with specific human genetic diseases. Family members SLC26A2, SLC26A3, and SLC26A4 have been recognized as the disease genes mutated in diastrophic dysplasia, congenital chloride diarrhea (CLD), and Pendred syndrome (PDS), respectively. DTDST is a particularly complex disorder. The gene encoding this molecule maps to chromosome
10 5q, and encodes two distinct transcripts due to alternative exon usage. In contrast to other sulfate transporters (*e.g.* Sat-1) anion movement by the DTDST protein is markedly inhibited by either extracellular chloride or bicarbonate. Impaired function of the DTDST gene product leads to undersulfation of proteoglycans and a complex family of recessively inherited osteochondrodysplasias (achondrogenesis type 1B,
15 atelosteogenesis type II, and diastrophic dysplasia) with clinical features including but not limited to, dwarfism, spinal deformation, and specific joint abnormalities. Interestingly, while epidemiological studies have shown that the disease occurs in most populations, it is particularly prevalent in Finland owing to an apparent founder effect. For these reasons, siRNAs directed against this class of genes (including but
20 not limited to SLC26A1-9, and related molecules) may be potentially helpful in both therapeutic and research venues.

Ion Exchangers

25 Intracellular pH regulatory mechanisms are critical for the maintenance of countless cellular processes. For instance, in muscle cells, contractile processes and metabolic reactions are influenced by pH. During periods of increased energy demands and ischemia, muscle cells produce large amounts of lactic acid that, without quick and efficient disposal, would lead to acidification of the sarcoplasm.

30 Several different transport mechanisms have evolved to maintain a relatively constant intracellular pH. The relative contribution of each of these processes varies with cell type, the metabolic requirements of the cell, and the local environmental conditions. Intracellular pH regulatory processes that have been characterized functionally include but are not limited to the Na^+/H^+ exchange, the $\text{Na}(\text{HCO}_3)_n$

cotransport, and the Na^+ -dependent and -independent Cl^- /base exchangers. As bicarbonate and CO_2 comprise the major pH buffer of biological fluids, sodium biocarbonate cotransporters (NBCs) are critical. Studies have shown that these molecules exist in numerous tissues including the kidney, brain, liver, cornea, heart, and lung, suggesting that NBCs play an important role in mediating HCO_3^- transport in both epithelial as well as nonepithelial cells.

Recent molecular cloning experiments have identified the existence of four NBC isoforms (NBC1, 2, 3 and 4) and two NBC-related proteins, AE4 and NCBE (Anion Exchanger 4 and Na-dependent Chloride-Bicarbonate Exchanger). The secondary structure analyses and hydropathy profile of this family predict them to be intrinsic membrane proteins with 12 putative transmembrane domains and several family members exhibit *N*-linked glycosylation sites, protein kinases A and C, casein kinase II, and ATP/GTP-binding consensus phosphorylation sites, as well as potential sites for myristylation and amidation. AE4 is a relatively recent addition to this family of proteins and shows between 30-48% homology with the other family members. When expressed in COS-7 cells and *Xenopus* oocytes AE4 exhibits sodium-independent and DIDS-insensitive anion exchanger activity. Exchangers have been shown to be responsible for a variety of human diseases. For instance, mutations in three genes of the anion transporter family (SLC) are believed to cause known hereditary diseases, including chondrodysplasia (SLC26A2, DTD), diarrhea (A3, down-regulated in adenoma/chloride-losing diarrhea protein: DRA/CLD), and goiter/deafness syndrome (A4, pendrin). Moreover, mutations in $\text{Na}^+/\text{HCO}_3^-$ cotransporters have also been associated with various human maladies. For these reasons, siRNAs directed against these sorts of genes (*e.g.* SLC4A4-10, and related genes) may be useful for therapeutic and research purposes.

Receptors Involved in Synaptic Transmission

In all vertebrates, fast inhibitory synaptic transmission is the result of the interaction between the neurotransmitters glycine (Gly) and γ -aminobutyric acid (GABA) and their respective receptors. The strychnine-sensitive glycine receptor is especially important in that it acts in the mammalian spinal cord and brain stem and has a well-established role in the regulation of locomotor behavior.

Glycine receptors display significant sequence homology to several other receptors including the nicotinic acetylcholine receptor, the aminobutyric acid receptor type A (GABA_AR), and the serotonin receptor type 3 (5-HT₃R) subunits. As members of the superfamily of ligand-gated ion channels, these polypeptides share common topological features. The glycine receptor is composed of two types of glycosylated integral membrane proteins (α 1- α 4 and β) arranged in a pentameric suprastructure. The alpha subunit encodes a large extracellular, N-terminal domain that carries the structural determinants essential for agonist and antagonist binding, followed by four transmembrane spanning regions (TM1-TM4), with TM2 playing the critical role of forming the inner wall of the chloride channel.

The density, location, and subunit composition of glycine neurotransmitter receptors changes over the course of development. It has been observed that the amount of GlyR gene translation (assessed by the injection of developing rat cerebral cortex mRNA into *Xenopus* oocytes) decreases with age, whereas that of GABA_ARs increases. In addition, the type and location of mRNAs coding for GlyR changes over the course of development. For instance in a study of the expression of alpha 1 and alpha 2 subunits in the rat, it was observed that (in embryonic periods E11-18) the mantle zone was scarce in the alpha 1 mRNA, but the germinal zone (matrix layer) at E11-14 expressed higher levels of the message. At postnatal day 0 (P0), the alpha 1 signals became manifested throughout the gray matter of the spinal cord. By contrast, the spinal tissues at P0 exhibited the highest levels of alpha-2 mRNA, which decreased with the postnatal development.

In both, man and mouse mutant lines, mutations of GlyR subunit genes result in hereditary motor disorders characterized by exaggerated startle responses and increased muscle tone. Pathological alleles of the *Glr1* gene are associated with the murine phenotypes *oscillator* (*spd^{ot}*) and *spasmodic* (*spd*). Similarly, a mutant allele of *Glr2* has been found to underly the molecular pathology of the *spastic* mouse (*spa*). Resembling the situation in the mouse, a variety of *GLRA1* mutant alleles have been shown to be associated with the human neurological disorder hyperekplexia or startle disease. For these reasons, siRNA directed against glycine receptors (GLRA1-3,

GLRB, and related molecules), glutamate receptors, GABA receptors, ATP receptors, and related neurotransmitter receptor molecules may be valuable therapeutic and research reagents.

5 **Proteases**

Kallikreins

One important class of proteases are the kallikreins, serine endopeptidases that split peptide substrates preferentially on the C-terminal side of internal arginyl and lysyl residues. Kallikreins are generally divided into two distinct groups, plasma kallikreins and tissue kallikreins. Tissue kallikreins represent a large group of enzymes that have substantial similarities at both the gene and protein level. The genes encoding this group are frequently found on a single chromosome, are organized in clusters, and are expressed in a broad range of tissues (*e.g.* pancreas, ovaries, breast). In contrast, the plasma form of the enzyme is encoded by a single gene (*e.g.* KLK3) that has been localized to chromosome 4q34-35 in humans. The gene encoding plasma kallikrein is expressed solely in the liver, contains 15 exons, and encodes a glycoprotein that is translated as a preprotein called prekallikrein.

Kallikreins are believed to play an important role in a host of physiological events. For instance, the immediate consequence of plasma prekallikrein activation is the cleavage of high molecular weight kininogen (HK) and the subsequent liberation of bradykinin, a nine amino acid vasoactive peptide that is an important mediator of inflammatory responses. Similarly, plasma kallikrein promotes single-chain urokinase activation and subsequent plasminogen activation, events that are critical to blood coagulation and wound healing.

Disruptions in the function of kallikreins have been implicated in a variety of pathological processes including imbalances in renal function and inflammatory processes. For these reasons, siRNAs directed against this class of genes (*e.g.* KLK1-15) may prove valuable in both research and therapeutic settings.

ADAM Proteins

The process of fertilization takes place in a series of discrete steps whereby the sperm interacts with, i) the cumulus cells and the hyaluronic acid extracellular matrix (ECM) in which they are embedded, ii) the egg's own ECM, called the *zona pellucida* (ZP), and iii) the egg plasma membrane. During the course of these interactions, the "acrosome reaction," the exocytosis of the acrosome vesicle on the head of the sperm, is induced, allowing the sperm to penetrate the ZP and gain access to the perivitelline space. This process exposes new portions of the sperm membrane, including the inner acrosomal membrane and the equatorial segment, regions of the sperm head that can participate in initial gamete membrane binding.

The interactions of the gamete plasma membranes appear to involve multiple ligands and receptors and are frequently compared to leukocyte-endothelial interactions. These interactions lead to a series of signal transduction events in the egg, known as collectively as egg activation and include the initiation of oscillations in intracellular calcium concentration, the exit from meiosis, the entry into the first embryonic mitosis, and the formation of a block to polyspermy via the release of ZP-modifying enzymes from the egg's cortical granules. Ultimately, sperm and egg not only adhere to each other but also go on to undergo membrane fusion, making one cell (the zygote) from two.

Studies on the process of sperm-egg interactions have identified a number of proteins that are crucial for fertilization. One class of proteins, called the ADAM family (A Disintegrin And Metalloprotease), has been found to be important in spermatogenesis and fertilization, as well as various developmental systems including myogenesis and neurogenesis. Members of the family contain a disintegrin and metalloprotease domain (and therefore have (potentially) both cell adhesion and protease activities), as well as cysteine-rich regions, epidermal growth factor (EGF)-like domains, a transmembrane region, and a cytoplasmic tail. Currently, the ADAM gene family has 29 members and constituents are widely distributed in many tissues including the brain, testis, epididymis, ovary, breast, placenta, liver, heart, lung, bone, and muscle.

- One of the best-studied members of the ADAM family is fertilin, a heterodimeric protein comprised of at least two subunits, fertilin alpha and fertilin beta. The fertilin beta gene (ADAM2) has been disrupted with a targeting gene construct corresponding to the exon encoding the fertilin beta disintegrin domain.
- 5 Sperm from males homozygous for disruptions in this region exhibit defects in multiple facets of sperm function including reduced levels of sperm transit from the uterus to the oviduct, reduced sperm-ZP binding, and reduced sperm-egg binding, all of which contribute to male infertility.
- 10 Recently, four new ADAM family members (ADAM 24-27) have been isolated. The deduced amino acid sequences show that all four contain the complete domain organization common to ADAM family members and Northern Blot analysis has shown all four to be specific to the testes. SiRNAs directed against this class of genes (*e.g.* ADAM2 and related proteins) may be useful as research tools and
- 15 therapeutics directed toward fertility and birth control.

Aminopeptidases

- Aminopeptidases are proteases that play critical roles in processes such as protein maturation, protein digestion in its terminal stage, regulation of hormone
- 20 levels, selective or homeostatic protein turnover, and plasmid stabilization. These enzymes generally have broad substrate specificity, occur in several forms and play a major role in physiological homeostasis. For instance, the effects of bradykinin, angiotensin converting enzyme (ACE), and other vasoactive molecules are muted by one of several peptidases that cleave the molecule at an internal position and eliminate
- 25 its ability to bind its cognate receptor (*e.g.* for bradykinin, the B2-receptor).

- Among the enzymes that can cleave bradykinin is the membrane bound aminopeptidase P, also referred to as aminoacylproline aminopeptidase, proline aminopeptidase; X-Pro aminopeptidase (eukaryote) and XPNPEP2. Aminopeptidase
- 30 P is an aminoacylproline aminopeptidase specific for NH₂-terminal Xaa-proline bonds. The enzyme i) is a mono-zinc-containing molecule that lacks any of the typical metal binding motifs found in other zinc metalloproteases, ii) has an active-site configuration similar to that of other members of the MG peptidase family, and iii) is

present in a variety of tissues including but not limited to the lung, kidney, brain, and intestine.

Aminopeptidases play an important role in a diverse set of human diseases.

- 5 Low plasma concentrations of aminopeptidase P are a potential predisposing factor for development of angio-oedema in patients treated with ACE inhibitors, and inhibitors of aminopeptidase P may act as cardioprotectors against other forms of illness including, but not limited to myocardial infarction. For these reasons, siRNAs directed against this family of proteins (including but not limited to XPNPEP1 and
10 related proteins) may be useful as research and therapeutic tools.

Serine Proteases

- One important class of proteases are the serine proteases. Serine proteases share a common catalytic triad of three amino acids in their active site (serine
15 (nucleophile), aspartate (electrophile), and histidine (base)) and can hydrolyze either esters or peptide bonds utilizing mechanisms of covalent catalysis and preferential binding of the transition state. Based on the position of their introns serine proteases have been classified into a minimum of four groups including those in which 1) the gene has no introns interrupting the exon coding for the catalytic triad (*e.g.* the
20 haptoglobin gene,); 2) each gene contains an intron just downstream from the codon for the histidine residue at the active site, a second intron downstream from the exon containing the aspartic acid residue of the active site and a third intron just upstream from the exon containing the serine of the active site (*e.g.* trypsinogen, chymotrypsinogen, kallikrein and proelastase); 3) the genes contain seven introns
25 interrupting the exons coding the catalytic region (*e.g.* complement factor B gene); and 4) the genes contain two introns resulting in a large exon that contains both the active site aspartic acid and serine residues (*e.g.* factor X, factor IX and protein C genes).

- 30 Cytotoxic lymphocytes (*e.g.* CD8(+) cytotoxic T cells and natural killer cells) form the major defense of higher organisms against virus-infected and transformed cells. A key function of these cells is to detect and eliminate potentially harmful cells by inducing them to undergo apoptosis. This is achieved through two principal pathways, both of which require direct but transient contact between the killer cell and

its target. The first pathway involves ligation of TNF receptor-like molecules such as Fas/CD95 to their cognate ligands, and results in mobilization of conventional, programmed cell-death pathways centered on activation of pro-apoptotic caspases. The second mechanism consists of a pathway whereby the toxic contents of a specialized class of secretory vesicles are introduced into the target cell. Studies over the last two decades have identified the toxic components as Granzymes, a family of serine proteases that are expressed exclusively by cytotoxic T lymphocytes and natural killer (NK) cells. These agents are stored in specialized lytic granules and enter the target cell via endocytosis. Like caspases, cysteine proteases that play an important role in apoptosis, granzymes can cleave proteins after acidic residues, especially aspartic acid, and induce apoptosis in the recipient cell.

Granzymes have been grouped into three subfamilies according to substrate specificity. Members of the granzyme family that have enzymatic activity similar to the serine protease chymotrypsin are encoded by a gene cluster termed the 'chymase locus'. Similarly, granzymes with trypsin-like specificities are encoded by the 'tryptase locus', and a third subfamily cleaves after unbranched hydrophobic residues, especially methionine, and are encoded by the 'Met-ase locus'. All granzymes are synthesized as zymogens and, after clipping of the leader peptide, obtain maximal enzymatic activity subsequent to the removal of an amino-terminal dipeptide.

Granzymes have been found to be important in a number of important biological functions including defense against intracellular pathogens, graft-versus-host reactions, the susceptibility to transplantable and spontaneous malignancies, lymphoid homeostasis, and the tendency toward auto-immune diseases. For these reasons, siRNAs directed against granzymes (*e.g.* GZMA, GZMB, GZMH, GZHK, GZMM) and related serine proteases may be useful research and therapeutic reagents.

Kinases

Protein Kinases (PKs) have been implicated in a number of biological processes. Kinase molecules play a central role in modulating cellular physiology and developmental decisions, and have been implicated in a large list of human maladies including cancer, diabetes, and others.

During the course of the last three decades, over a hundred distinct protein kinases have been identified, all with presumed specific cellular functions. A few of these enzymes have been isolated to sufficient purity to perform *in vitro* studies, but most remain intractable due to the low abundance of these molecules in the cell. To counter this technical difficulty, a number of protein kinases have been isolated by molecular cloning strategies that utilize the conserved sequences of the catalytic domain to isolate closely related homologs. Alternatively, some kinases have been purified (and subsequently studied) based on their interactions with other molecules.

p58 is a member of the p34cdc2-related supergene family and contains a large domain that is highly homologous to the cell division control kinase, cdc2. This new cell division control-related protein kinase was originally identified as a component of semipurified galactosyltransferase; thus, it has been denoted galactosyltransferase-associated protein kinase (GTA-kinase). GTA-kinase has been found to be expressed in both adult and embryonic tissues and is known to phosphorylate a number of substrates, including histone H1, and casein. Interestingly enough, over expression of this molecule in CHO cells has shown that elevated levels of p58 result in a prolonged late telophase and an early G1 phase, thus hinting of an important role for GTA-kinase in cell cycle regulation.

Cyclin Dependent Kinases

The cyclin-dependent kinases (Cdks) are a family of highly conserved serine/threonine kinases that mediate many of the cell cycle transitions that occur during duplication. Each of these Cdk catalytic subunits associates with a specific subset of regulatory subunits, termed cyclins, to produce a distinct Cdk-cyclin kinase complex that, in general, functions to execute a unique cell cycle event.

Activation of the Cdk-cyclin kinases during cellular transitions is controlled by a variety of regulatory mechanisms. For the Cdc2-cyclin B complex, inhibition of kinase activity during S phase and G₂ is accomplished by phosphorylation of two Cdc2 residues, Thr¹⁴ and Tyr¹⁵, which are positioned within the ATP-binding cleft. Phosphorylation of Thr¹⁴ and/or Tyr¹⁵ suppresses the catalytic activity of the molecule by disrupting the orientation of the ATP present within this cleft. In contrast, the abrupt dephosphorylation of these residues by the Cdc25 phosphatase results in the

rapid activation of Cdc2-cyclin B kinase activity and subsequent downstream mitotic events. While the exact details of this pathway have yet to be elucidated, it has been proposed that Thr¹⁴/Tyr¹⁵ phosphorylation functions to permit a cell to attain a critical concentration of inactive Cdk-cyclin complexes, which, upon activation, induces a rapid and complete cell cycle transition. Furthermore, there is evidence in mammalian cells that Thr¹⁴/Tyr¹⁵ phosphorylation also functions to delay Cdk activation after DNA damage.

The *Schizosaccharomyces pombe weel* gene product was the first kinase identified that is capable of phosphorylating Tyr¹⁵ in Cdc2. Homologs of the Wee1 kinase have been subsequently identified and biochemically characterized from a wide range of species including human, mouse, frog, *Saccharomyces cerevisiae*, and *Drosophila*. In vertebrate systems, where Thr¹⁴ in Cdc2 is also phosphorylated, the Wee1 kinase was capable of phosphorylating Cdc2 on Tyr¹⁵, but not Thr¹⁴, indicating that another kinase was responsible for Thr¹⁴ phosphorylation. This gene, Myt1 kinase, was recently isolated from the membrane fractions of *Xenopus* egg extracts and has been shown to be capable of phosphorylating Thr¹⁴ and, to a lesser extent, Tyr¹⁵ in Cdc2. A human Myt1 homolog displaying similar properties has been isolated, as well as a non-membrane-associated molecule with Thr¹⁴ kinase activity.

In the past decade it has been shown that cancer can originate from overexpression of positive regulators, such as cyclins, or from underexpression of negative regulators (e.g. p16 (INK4a), p15 (INK4b), p21 (Cip1)). Inhibitors such as Myt1 are the focus of much cancer research because they are capable of controlling cell cycle proliferation, now considered the Holy Grail for cancer treatment. For these reasons, siRNA directed against kinases and kinase inhibitors including but not limited to ABL1, ABL2, ACK1, ALK, AXL, BLK, BMX, BTK, C20orf64, CSF1R, SCK, DDR1, DDR2, DKFZp761P1010, EGFR, EPHA1, EPHA2, EPHA3, EPHA4, EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, EPHB6, ERBB2, ERBB3, ERBB4, FER, FES, FGFR1, FGFR2, FGFR3, FGFR4, FGR, FLT1, FLT3, FLT4, FRK, FYN, HCK, IGF1R, INSR, ITK, JAK1, JAK2, JAK3, KDR, KIAA1079, KIT, LCK, LTK, LYN, MATK, MERTK, MET, MST1R, MUSK, NTRK1, NTRK2, NTRK3, PDGFRA, PDGFRB, PTK2, PTK2B, PTK6, PTK7, PTK9, PTK9L, RET, ROR1, ROR2, ROS1, RYK, SRC, SYK, TEC, TEK, TIE, TNK1, TXK, TYK2,

TYRO3, YES1, and related proteins, may be useful for research and therapeutic purposes.

G Protein Coupled Receptors

5

One important class of genes to which siRNAs can be directed are G-protein coupled receptors (GPCRs). GPCRs constitute a superfamily of seven transmembrane spanning proteins that respond to a diverse array of sensory and chemical stimuli, such as light, odor, taste, pheromones, hormones and neurotransmitters. GPCRs play a central role in cell proliferation, differentiation, and have been implicated in the etiology of disease.

The mechanism by which G protein-coupled receptors translate extracellular signals into cellular changes was initially envisioned as a simple linear model: activation of the receptor by agonist binding leads to dissociation of the heterotrimeric GTP-binding G protein (Gs, Gi, or Gq) into its alpha and beta/gamma subunits, both of which can activate or inhibit various downstream effector molecules. More specifically, activation of the GPCR induces a conformational change in the G α subunit, causing GDP to be released and GTP to be bound in its place. The G α and G $\beta\gamma$ subunits then dissociate from the receptor and interact with a variety of effector molecules. For instance in the case of the Gs family, the primary function is to stimulate the intracellular messenger adenylate cyclase (AC), which catalyzes the conversion of cytoplasmic ATP into the secondary messenger cyclic AMP (cAMP). In contrast, the Gi family inhibits this pathway and the Gq family activates phospholipases C (PLC), which cleaves phosphatidylinositol 4,5, bisphosphate (PIP2) to generate inositol-1,4,5-phosphate (IP3) and diacylglycerol (DAG).

More recently, studies have shown that the functions of GPCRs are not limited to their actions on G-proteins and that considerable cross-talk exists between this diverse group of receptor molecules and a second class of membrane bound proteins, the receptor tyrosine kinases (RTKs). A number of GPCRs such as endothelin-1, thrombin, bombesin, and dopamine receptors can activate MAPKs, a downstream effector of the RTK/Ras pathway. Interestingly, the interaction between these two families is not unidirectional and RTKs can also modulate the activity of signaling

pathways traditionally thought to be controlled exclusively by ligands that couple to GPCRs. For instance, EGF, which normally activates the MAPK cascade via the EGF receptor can stimulate adenylate cyclase activity by activating G α s.

- 5 There are dozens of members of the G Protein-Coupled Receptor family that have emerged as prominent drug targets in the last decade. One non-limiting list of potential GPCR-siRNA targets is as follows:

CMKLR1

- 10 CML1/ CMKLR1 (Accession No. Q99788) is a member of the chemokine receptor family of GPCRs that may play a role in a number of diseases including those involved in inflammation and immunological responses (*e.g.* asthma, arthritis). For this reason, siRNA directed against this protein may prove to be important therapeutic reagents.

15

- Studies of juvenile-onset neuronal ceroid lipofuscinosis (JNCL, Batten disease), the most common form of childhood encephalopathy that is characterized by progressive neural degeneration, show that it is brought on by mutations in a novel lysosomal membrane protein (CLN3). In addition to being implicated in JNCL,
- 20 CLN3 (GPCR-like protein, Accession No. A57219) expression studies have shown that the CLN3 mRNA and protein are highly over-expressed in a number of cancers (*e.g.* glioblastomas, neuroblastomas, as well as cancers of the prostate, ovaries, breast, and colon) suggesting a possible contribution of this gene to tumor growth. For this reason, siRNA directed against this protein may prove to be important therapeutic
- 25 reagents.

CLACR

- The calcitonin receptor (CTR/ CALCR, Accession No. NM_001742) belongs to “family B” of GPCRs which typically recognized regulatory peptides such as
- 30 parathyroid hormone, secretin, glucagons and vasoactive intestinal polypeptide. Although the CT receptor typically binds to calcitonin (CT), a 32 amino acid peptide hormone produced primarily by the thyroid, association of the receptor with RAMP (Receptor Activity Modulating Protein) enables it to readily bind other members of

the calcitonin peptide family including amylin (AMY) and other CT gene-related peptides (*e.g.* α CGRP and β CGRP). While the primary function of the calcitonin receptor pertains to regulating osteoclast mediated bone resorption and enhanced Ca^{+2} excretion by the kidney, recent studies have shown that CT and CTRs may play an important role in a variety of processes as wide ranging as embryonic/foetal development and sperm function/physiology. In addition, studies have shown that patients with particular CTR genotypes may be at higher risk to lose bone mass and that this GPCR may contribute to the formation of calcium oxalate urinary stones. For this reason, siRNA directed against CTR may be useful as therapeutic reagents.

OXTR

The human oxytocin receptor (OTR, OXTR) is a 389 amino acid polypeptide that exhibits the seven transmembrane domain structure and belongs to the Class-I (rhodopsin-type) family of G-protein coupled receptors. OTR is expressed in a wide variety of tissues throughout development and mediates physiological changes through G(q) proteins and phospholipase C-beta. Studies on the functions of oxytocin and the oxytocin receptor have revealed a broad list of duties. OT and OTR play a role in a host of sexual, maternal and social behaviors that include egg-laying, birth, milk-letdown, feeding, grooming, memory and learning. In addition, it has been hypothesized that abnormalities in the functionality of oxytocin-OTR receptor-ligand system can lead to a host of irregularities including compulsive behavior, eating disorders (such as anorexia), depression, and various forms of neurodegenerative diseases. For these reasons, siRNA directed against this gene (NM_000916) may play an important role in combating OTR-associated illnesses.

EDG GPCRs

Lysophosphatidic acid and other lipid-based hormones/growth factors induce their effects by activating signaling pathways through the G-protein coupled receptors (GPCRs) and have been observed to play important roles in a number of human diseases including cancer, asthma, and vascular pathologies. For instance, during studies of immunoglobulin A nephropathy (IgAN), researchers have observed an enhanced expression of EDG5 (NP_004221) suggesting a contribution of this gene product in the development of IgAN. For that reasons, siRNA directed against Edg5

(NM_004230), Edg4 (NM_004720), Edg7 (NM_012152) and related genes may play an important role in combating human disease.

Genes Involved in Cholesterol Signaling and Biosynthesis

5 Studies on model genetic organisms such as *Drosophila* and *C. elegans* have led to the identification of a plethora of genes that are essential for early development. Mutational analysis and ectopic expression studies have allowed many of these genes to be grouped into discrete signal transduction pathways and have shown that these elements play critical roles in pattern formation and cell differentiation. Disruption of
10 one or more of these genes during early stages of development frequently leads to birth defects whereas alteration of gene function at later stages in life can result in tumorigenesis.

 One critical set of interactions known to exist in both invertebrates and
15 vertebrates is the Sonic Hedgehog-Patched-Gli pathway. Originally documented as a *Drosophila* segmentation mutant, several labs have recently identified human and mouse orthologs of many of the pathway members and have successfully related disruptions in these genes to known diseases. Pathway activation is initiated with the secretion of Sonic hedgehog. There are three closely related members of the Shh
20 family (Sonic hedgehog, Desert, and Indian) with Shh being the most widely expressed form of the group. The Shh gene product is secreted as a small pro-signal molecule. To successfully initiate its developmental role, Shh is first cleaved, whereupon the N-terminal truncated fragment is covalently modified with cholesterol. The addition of the sterol moiety promotes the interaction between Shh and its
25 cognate membrane bound receptor, Patched (Ptch). There are at least two isoforms of the Patched gene, Ptch1 and Ptch2. Both isoforms contain a sterol-sensing domain (SSD); a roughly 180 amino acid cluster that is found in at least seven different classes of molecules including those involved in cholesterol biosynthesis, vesicular traffic, signal transduction, cholesterol transport, and sterol homeostasis. In the
30 absence of Shh, the Patched protein is a negative regulator of the pathway. In contrast, binding of Shh-cholesterol to the Patched receptor releases the negative inhibition which that molecule enforces on a G-protein coupled receptor known as Smoothened. Subsequent activation of Smoothened (directly or indirectly) leads to the triggering of a trio of transcription factors that belong to the Gli family. All three factors are

relatively large, contain a characteristic C2-H2 zinc-finger pentamer, and recognize one of two consensus sequences (SEQ. ID NO. 0463 GACCACCCA or SEQ. ID NO. 0464 GAACCACCCA). In the absence of Shh, Gli proteins are cleaved by the proteasome and the C-terminally truncated fragment translocates to the nucleus and
5 acts as a dominant transcription repressor. In the presence of Shh-cholesterol, Gli repressor formation is inhibited and full-length Gli functions as a transcriptional activator.

Shh and other members of the Shh-PTCH-Gli pathway are expressed in a
10 broad range of tissues (*e.g.* the notochord, the floorplate of the neural tube, the brain, and the gut) at early stages in development. Not surprisingly, mutations that lead to altered protein expression or function have been shown to induce developmental abnormalities. Defects in the human Shh gene have been shown to cause holoprosencephaly, a midline defect that manifests itself as cleft lip or palate, CNS
15 septation, and a wide range of other phenotypes. Interestingly, defects in cholesterol biosynthesis generate similar Shh-like disorders (*e.g.* Smith-Lemli-Opitz syndrome) suggesting that cholesterol modification of the Shh gene product is crucial for pathway function. Both the Patched and Smoothed genes have also been shown to be clinically relevant with Smoothed now being recognized as an oncogene that,
20 like PTCH-1 and PTCH-2, is believed to be the causative agent of several forms of adult tumors. For these reasons, siRNA directed against Smoothed (SMO, NM_005631), Patched (PTCH, nm_000264), and additional genes that participate in cholesterol signaling, biosynthesis, and degradation, have potentially useful research and therapeutic applications.

25

Targeted Pathways.

In addition to targeting siRNA against one or more members of a family of proteins, siRNA can be directed against members of a pathway. Thus, for instance, siRNA can be directed against members of a signal transduction pathway (*e.g.* the
30 insulin pathway, including AKT1-3, CBL, CBLB, EIF4EBP1, FOXO1A, FOXO3A, FRAP1, GSK3A, GSK3B, IGF1, IGF1R, INPP5D, INSR, IRS1, MLLT7, PDPK1, PIK3CA, PIK3CB, PIK3R1, PIK3R2, PPP2R2B, PTEN, RPS6, RPS6KA1, RPX6KA3, SGK, TSC1, TSC2, AND XPO1), an apoptotic pathway (CASP3,6,7,8,9, DSH1/2, P110, P85, PDK1/2, CATENIN, HSP90, CDC37, P23, BAD, BCLXL,

BCL2, SMAC, and others), pathways, involved in DNA damage, cell cycle, and other physiological (p53,MDM2, CHK1/2, BRCA1/2, ATM, ATR, P15^{INK4}, P27, P21, SKP2, CDC25C/A, 14-3-3, PLK, RB, CDK4, GLUT4, Inos, Mtor, FKBP, PPAR, RXR, ER). Similarly, genes involved in immune system function including TNFR1, IL-IR, IRAK1/2, TRAF2, TRAF6, TRADD, FADD, IKK ϵ , IKK γ , IKK β , IKK α , I κ B α , I κ B β , p50, p65, Rac, RhoA, Cdc42, ROCK, Pak1/2/3/4/5/6, cIAP, HDAC1/2, CBP, β -TrCP, Rip2/4, and others are also important targets for the siRNAs described in this document and may be useful in treating immune system disorders. Genes involved in apoptosis, such as Dsh1/2,PTEN, P110 (pan), P85, PDK1/2, Akt1, Akt2, Akt (pan), p70^{S6K}, GSK3 β , PP2A (cat), β -catenin, HSP90, Cdc37/p50, P23, Bad, BclxL, Bcl2, Smac/Diablo, and Ask1 are potentially useful in the treatment of diseases that involve defects in programmed cell death (*e.g.* cancer), while siRNA agents directed against p53, MDM2, Chk1/2, BRCA1/2, ATM, ATR, p15^{INK4}, P27, P21, Skp2, Cdc25C/A, 14-3-3 σ/ϵ , PLK, Rb, Cdk4, Glut4, iNOS, mTOR, FKBP, PPAR γ , RXR α , ER α and related genes may play a critical role in combating diseases associated with disruptions in DNA repair, and cell cycle abnormalities.

Tables VI -Table X below provide examples of useful pools for inhibiting different genes in the human insulin pathway and tyrosine kinase pathways, proteins involved in the cell cycle, the production of nuclear receptors, and other genes. These particular pools are particularly useful in humans, but would be useful in any species that generates an appropriately homologous mRNA. Further, within each of the listed pools any one sequence maybe used independently but preferably at least two of the listed sequences, more preferably at least three, and most preferably all of the listed sequences for a given gene is present.

Table VI

Gene Name	Acc#	GI	L.L.	Duplex #	Sequence	SEQ. ID NO.
AKT1	NM 005163	4885060	207	D-003000-05	GACAAGGACGGGCACATTA	465
AKT1	NM 005163	4885060	207	D-003000-06	GGACAAGGACGGGCACATT	466
AKT1	NM 005163	4885060	207	D-003000-07	GCTACTTCCTCCTCAAGAA	467
AKT1	NM 005163	4885060	207	D-003000-08	GACCGCCTCTGCTTTGTCA	468
AKT2						
AKT2	NM 001626	6715585	208	D-003001-05	GTACTTCGATGATGAATTT	469
AKT2	NM 001626	6715585	208	D-003001-06	GCAAAGAGGGGCATCAGTGA	470
AKT2	NM 001626	6715585	208	D-003001-07	GGGCTAAAGTGACCATGAA	471

AKT2	NM 001626	6715585	208	D-003001-08	GCAGAATGCCAGCTGATGA	472
AKT3						
AKT3	NM 005465	32307164	10000	D-003002-05	GGAGTAACTGGCAAGATG	473
AKT3	NM 005465	32307164	10000	D-003002-06	GACATTAAATTTCTCGAA	474
AKT3	NM 005465	32307164	10000	D-003002-07	GACCAAAGCCAAACACATT	475
AKT3	NM 005465	32307164	10000	D-003002-08	GAGGAGAGAATGAATTGTA	476
CBL						
CBL	NM 005188	4885116	867	D-003003-05	GGAGACACATTTCTGGATTA	477
CBL	NM 005188	4885116	867	D-003003-06	GATCTGACCTGCAATGATT	478
CBL	NM 005188	4885116	867	D-003003-07	GACAATCCCTCACAATAAA	479
CBL	NM 005188	4885116	867	D-003003-08	CCAGAAAGCTTTGGTCATT	480
CBLB						
CBLB	NM 170662	29366807	868	D-003004-05	GACCATACCTCATAACAAG	481
CBLB	NM 170662	29366807	868	D-003004-06	TGAAAGACCTCCACCAATC	482
CBLB	NM 170662	29366807	868	D-003004-07	GATGAAGGCTCCAGGTGTT	483
CBLB	NM 170662	29366807	868	D-003004-08	TATCAGCATTTACGACTTA	484
EIF4EBP1						
EIF4EBP1	NM 004095	20070179	1978	D-003005-05	GCAATAGCCCAGAAGATAA	485
EIF4EBP1	NM 004095	20070179	1978	D-003005-06	CGCAATAGCCCAGAAGATA	486
EIF4EBP1	NM 004095	20070179	1978	D-003005-07	GAGATGGACATTTAAAGCA	487
EIF4EBP1	NM 004095	20070179	1978	D-003005-08	CAATAGCCCAGAAGATAAG	488
FOXO1A						
FOXO1A	NM 002015	9257221	2308	D-003006-05	CCAGGCATCTCATAACAAA	489
FOXO1A	NM 002015	9257221	2308	D-003006-06	CCAGATGCCTATACAAACA	490
FOXO1A	NM 002015	9257221	2308	D-003006-07	GGAGGTATGAGTCAGTATA	491
FOXO1A	NM 002015	9257221	2308	D-003006-08	GAGGTATGAGTCAGTATAA	492
FOXO3A						
FOXO3A	NM 001455	4503738	2309	D-003007-01	CAATAGCAACAAGTATACC	493
FOXO3A	NM 001455	4503738	2309	D-003007-02	TGAAGTCCAGGACGATGAT	494
FOXO3A	NM 001455	4503738	2309	D-003007-03	TGTCACACTATGGTAACCA	495
FOXO3A	NM 001455	4503738	2309	D-003007-04	TGTTCAATGGGAGCTTGGA	496
FRAP1						
FRAP1	NM 004958	19924298	2475	D-003008-05	GAGAAGAAATGGAAGAAAT	497
FRAP1	NM 004958	19924298	2475	D-003008-06	CCAAAGTGCTGCAGTACTA	498
FRAP1	NM 004958	19924298	2475	D-003008-07	GAGCATGCCGTCAATAATA	499
FRAP1	NM 004958	19924298	2475	D-003008-08	GGTCTGAAGTGAATGAAGA	500
GSK3A						
GSK3A	NM 019884	11995473	2931	D-003009-05	GGACAAAGGTGTTCAAATC	501
GSK3A	NM 019884	11995473	2931	D-003009-06	GAACCCAGCTGCCTAACAA	502
GSK3A	NM 019884	11995473	2931	D-003009-07	GCGCACAGCTTCTTTGATG	503
GSK3A	NM 019884	11995473	2931	D-003009-08	GCTCTAGCCTGCTGGAGTA	504
GSK3B						
GSK3B	NM 002093	21361339	2932	D-003010-05	GAAGAAAGATGAGGTCTAT	505
GSK3B	NM 002093	21361339	2932	D-003010-06	GGACCCAAATGTCAAATA	506
GSK3B	NM 002093	21361339	2932	D-003010-07	GAAATGAACCCAACTACA	507
GSK3B	NM 002093	21361339	2932	D-003010-08	GATGAGGTCTATCTTAATC	508
IGF1						
IGF1	NM 000618			D-003011-05	GGAAGTACATTTGAAGAAC	509
IGF1	NM 000618			D-003011-06	AGAAGGAAGTACATTTGAA	510
IGF1	NM 000618			D-003011-07	CCTCAAGCCTGCCAAGTCA	511
IGF1	NM 000618			D-003011-08	GGTGGATGCTCTTCAGTTC	512
IGF1R						
IGF1R	NM 000875	11068002	3480	D-003012-05	CAACGAAGCTTCTGTGATG	513
IGF1R	NM 000875	11068002	3480	D-003012-06	GGCCAGAAATGGAGAATAA	514
IGF1R	NM 000875	11068002	3480	D-003012-07	GAAGCACCTTTAAGAATG	515
IGF1R	NM 000875	11068002	3480	D-003012-08	GCAGACACCTACAACATCA	516

INPP5D						
INPP5D	NM 005541	5031798	3635	D-003013-05	GGAATTGCGTTTACACTTA	517
INPP5D	NM 005541	5031798	3635	D-003013-06	GGAAACTGATCATTAAAGAA	518
INPP5D	NM 005541	5031798	3635	D-003013-07	CGACAGGGATGAAGTACAA	519
INPP5D	NM 005541	5031798	3635	D-003013-08	AAACGCAGCTGCCCATCTA	520
INSR						
INSR	NM 000208	4557883	3643	D-003014-05	GGAAGACGTTTGAGGATTA	521
INSR	NM 000208	4557883	3643	D-003014-06	GAACAAGGCTCCCGAGAGT	522
INSR	NM 000208	4557883	3643	D-003014-07	GGAGAGACCTTGAAATTG	523
INSR	NM 000208	4557883	3643	D-003014-08	GGACGGAACCCACCTATTT	524
IRS1						
IRS1	NM 005544	5031804	3667	D-003015-05	AAAGAGGTCTGGCAAGTGA	525
IRS1	NM 005544	5031804	3667	D-003015-06	GAACCTGATTGGTATCTAC	526
IRS1	NM 005544	5031804	3667	D-003015-07	CCACGGCGATCTAGTGCTT	527
IRS1	NM 005544	5031804	3667	D-003015-08	GTCAGTCTGTCGTCCAGTA	528
MLLT7						
MLLT7	NM 005938	5174578	4303	D-003016-05	GGACTGGACTTCAACTTTG	529
MLLT7	NM 005938	5174578	4303	D-003016-06	CCACGAAGCAGTTCAAATG	530
MLLT7	NM 005938	5174578	4303	D-003016-07	GAGAAGCGACTGACACTTG	531
MLLT7	NM 005938	5174578	4303	D-003016-08	GACCAGAGATCGCTAACCA	532
PDPK1						
PDPK1	NM 002613	4505694	5170	D-003017-05	CAAGAGACCTCGTGGAGAA	533
PDPK1	NM 002613	4505694	5170	D-003017-06	GACCAGAGGCCAAGAATTT	534
PDPK1	NM 002613	4505694	5170	D-003017-07	GGAAACGAGTATCTTATAT	535
PDPK1	NM 002613	4505694	5170	D-003017-08	GAGAAGCGACATATCATAA	536
PIK3CA						
PIK3CA	NM 006218	5453891	5290	D-003018-05	GCTATCATCTGAACAATTA	537
PIK3CA	NM 006218	5453891	5290	D-003018-06	GGATAGAGGCCAAATAATA	538
PIK3CA	NM 006218	5453891	5290	D-003018-07	GGACAACTGTTTCATATAG	539
PIK3CA	NM 006218	5453891	5290	D-003018-08	GCCAGTACCTCATGGATTA	540
PIK3CB						
PIK3CB	NM 006219	5453893	5291	D-003019-05	CGACAAGACTGCCGAGAGA	541
PIK3CB	NM 006219	5453893	5291	D-003019-06	TCAAGTGTCTCCTAATATG	542
PIK3CB	NM 006219	5453893	5291	D-003019-07	GGATTCAAGTTGGAGTGATT	543
PIK3CB	NM 006219	5453893	5291	D-003019-08	TTTCAAGTGTCTCCTAATA	544
PIK3R1						
PIK3R1	NM 181504	32455251	5295	D-003020-05	GGAAATATGGCTTCTCTGA	545
PIK3R1	NM 181504	32455251	5295	D-003020-06	GAAAGACGAGAGACCAATA	546
PIK3R1	NM 181504	32455251	5295	D-003020-07	GTAAAGCATTGTGTCATAA	547
PIK3R1	NM 181504	32455251	5295	D-003020-08	GGATCAAGTTGTCAAAGAA	548
PIK3R2						
PIK3R2	NM 005027	4826907	5296	D-003021-05	GGAAAGGCGGGAACAATAA	549
PIK3R2	NM 005027	4826907	5296	D-003021-06	GATGAAGCGTACTGCAATT	550
PIK3R2	NM 005027	4826907	5296	D-003021-07	GGACAGCGAATCTCACTAC	551
PIK3R2	NM 005027	4826907	5296	D-003021-08	GCAAGATCCGAGACCAGTA	552
PPP2R2B						
PPP2R2B	NM 004576	4758953	5521	D-003022-05	GAATGCAGCTTACTTTCTT	553
PPP2R2B	NM 004576	4758953	5521	D-003022-06	GACCGAAGCTGACATTATC	554
PPP2R2B	NM 004576	4758953	5521	D-003022-07	TCGATTACCTGAAGAGTTT	555
PPP2R2B	NM 004576	4758953	5521	D-003022-08	CCTGAAGAGTTTAGAAATA	556
PTEN						
PTEN	NM 000314	4506248	5728	D-003023-05	GTGAAGATCTTGACCAATG	557
PTEN	NM 000314	4506248	5728	D-003023-06	GATCAGCATACACAAATTA	558
PTEN	NM 000314	4506248	5728	D-003023-07	GGCGCTATGTGTATTATTA	559
PTEN	NM 000314	4506248	5728	D-003023-08	GTATAGAGCGTGCAGATAA	560
RPS6						

RPS6	NM_001010	17158043	6194	D-003024-05	GCCAGAACTCATTGAAGT	561
RPS6	NM_001010	17158043	6194	D-003024-06	GGATATTCCTGGACTGACT	562
RPS6	NM_001010	17158043	6194	D-003024-07	CCAAGGAGAACTGGAGAAA	563
RPS6	NM_001010	17158043	6194	D-003024-08	GCGTATGGCCACAGAAGTT	564
RPS6KA1						
RPS6KA1	NM_002953	20149546	6195	D-003025-05	GATGACACCTTCTACTTTG	565
RPS6KA1	NM_002953	20149546	6195	D-003025-06	GAGAATGGGCTCCTCATGA	566
RPS6KA1	NM_002953	20149546	6195	D-003025-07	CAAGCGGGATCCTTCAGAA	567
RPS6KA1	NM_002953	20149546	6195	D-003025-08	CCACCGGCCTGATGGAAGA	568
RPS6KA3						
RPS6KA3	NM_004586	4759049	6197	D-003026-05	GAAGGGAAGTTGTATCTTA	569
RPS6KA3	NM_004586	4759049	6197	D-003026-06	GAAAGTATGTGTATGTAGT	570
RPS6KA3	NM_004586	4759049	6197	D-003026-07	GGACAGCATCCAAACATTA	571
RPS6KA3	NM_004586	4759049	6197	D-003026-08	GGAGGTGAATTGCTGGATA	572
SGK						
SGK	NM_005627	5032090	6446	D-003027-01	TTAATGGTGGAGAGTTGTT	573
SGK	NM_005627	5032090	6446	D-003027-04	ATTAAGTGGGATGATCTCA	574
SGK	NM_005627	25168262	6446	D-003027-05	GAAGAAAGCAATCCTGAAA	575
SGK	NM_005627	25168262	6446	D-003027-06	AAACACAGCTGAAATGTAC	576
TSC1						
TSC1	NM_000368	24475626	7248	D-003028-05	GAAGATGGCTATTCTGTGT	577
TSC1	NM_000368	24475626	7248	D-003028-06	TATGAAGGCTCGAGAGTTA	578
TSC1	NM_000368	24475626	7248	D-003028-07	CGACACGGCTGATAACTGA	579
TSC1	NM_000368	24475626	7248	D-003028-08	CGGCTGATGTTGTTAAATA	580
TSC2						
TSC2	NM_000548	10938006	7249	D-003029-05	GCATTAATCTCTTACCATA	581
TSC2	NM_000548	10938006	7249	D-003029-06	CCAATGTCCTCTTGTCTTT	582
TSC2	NM_000548	10938006	7249	D-003029-07	GGAGACACATCACCTACTT	583
TSC2	NM_000548	10938006	7249	D-003029-08	TCACCAGGCTCATCAAGAA	584
XPO1						
XPO1	NM_003400	8051634	7514	D-003030-05	GAAAGTCTCTGTCAAAAATA	585
XPO1	NM_003400	8051634	7514	D-003030-06	GCAATAGGCTCCATTAGTG	586
XPO1	NM_003400	8051634	7514	D-003030-07	GGAACATGATCAACTTATA	587
XPO1	NM_003400	8051634	7514	D-003030-08	GGATACAGATTCCATAAAT	588

Table VII

Gene Name	Acc#	GI	L.L	Duplex #	Sequence	SEQ. ID NO.
ABL1						
ABL1	NM_007313	6382057	25	D-003100-05	GGAAATCAGTGACATAGTG	589
ABL1	NM_007313	6382057	25	D-003100-06	GGTCCACACTGCAATGTTT	590
ABL1	NM_007313	6382057	25	D-003100-07	GAAGGAAATCAGTGACATA	591
ABL1	NM_007313	6382057	25	D-003100-08	TCACTGAGTTCATGACCTA	592
ABL2						
ABL2	NM_007314	6382061	27	D-003101-05	GAAATGGAGCGAACAGATA	593
ABL2	NM_007314	6382061	27	D-003101-06	GAGCCAAATTTCTATTAA	594
ABL2	NM_007314	6382061	27	D-003101-07	GTAATAAGCCTACAGTCTA	595
ABL2	NM_007314	6382061	27	D-003101-08	GGAGTGAAGTTCGCTCTAA	596
ACK1						
ACK1	NM_005781	8922074	10188	D-003102-05	AAACGCAAGTCGTGGATGA	597
ACK1	NM_005781	8922074	10188	D-003102-06	GCAAGTCGTGGATGAGTAA	598
ACK1	NM_005781	8922074	10188	D-003102-07	GAGCACTACCTCAGAATGA	599

ACK1	NM_005781	8922074	10188	D-003102-08	TCAGCAGCACCCACTATTA	600
ALK						
ALK	NM_004304	29029631	238	D-003103-05	GACAAGATCCTGCAGAATA	601
ALK	NM_004304	29029631	238	D-003103-06	GGAAGAGTCTGGCAGTTGA	602
ALK	NM_004304	29029631	238	D-003103-07	GCACGTGGCTCGGGACATT	603
ALK	NM_004304	29029631	238	D-003103-08	GAACTGCAGTGAAGGAACA	604
AXL						
AXL	NM_021913	21536465	558	D-003104-05	GGTCAGAGCTGGAGGATTT	605
AXL	NM_021913	21536465	558	D-003104-06	GAAAGAAGGAGACCCGTTA	606
AXL	NM_021913	21536465	558	D-003104-07	CCAAGAAGATCTACAATGG	607
AXL	NM_021913	21536465	558	D-003104-08	GGAAGTGCATGCTGAATGA	608
BLK						
BLK	NM_001715	4502412	640	D-003105-05	GAGGATGCCTGCTGGATTT	609
BLK	NM_001715	4502412	640	D-003105-06	ACATGAAGGTGGCCATTAA	610
BLK	NM_001715	4502412	640	D-003105-07	GGTCAGCGCCCAAGACAAG	611
BLK	NM_001715	4502412	640	D-003105-08	GAAACTCGGGTCTGGACAA	612
BMX						
BMX	NM_001721	21359831	660	D-003106-05	AAACAAACCTTTCCTACTA	613
BMX	NM_001721	21359831	660	D-003106-06	GAAGGAGCATTTATGGTTA	614
BMX	NM_001721	21359831	660	D-003106-07	GAGAAGAGATTACCTTGT	615
BMX	NM_001721	21359831	660	D-003106-08	GTAAGGCTGTGAATGATAA	616
BTK						
BTK	NM_000061	4557376	695	D-003107-05	GAACAGGAATGGAAGCTTA	617
BTK	NM_000061	4557376	695	D-003107-06	GCTATGGGCTGCCAAATTT	618
BTK	NM_000061	4557376	695	D-003107-07	GAAAGCAACTTACCATGGT	619
BTK	NM_000061	4557376	695	D-003107-08	GGTAAACGATCAAGGAGTT	620
C20orf64						
C20orf64	NM_033550	19923655	11285	D-003108-05	CAACTTAGCCAAGACAATT	621
C20orf64	NM_033550	19923655	11285	D-003108-06	GAAATTGAAGGCTCAGTGA	622
C20orf64	NM_033550	19923655	11285	D-003108-07	TGGAACAGCTGAACATTGT	623
C20orf64	NM_033550	19923655	11285	D-003108-08	GCTTCCAAGTCTTATATA	624
CSF1R						
CSF1R	NM_005211	27262658	1436	D-003109-05	GGAGAGCTCTGAGCTTTGA	625
CSF1R	NM_005211	27262658	1436	D-003109-06	CAACAACGCTACCTTCCAA	626
CSF1R	NM_005211	27262658	1436	D-003109-07	CCACGCAGCTGCCTTACAA	627
CSF1R	NM_005211	27262658	1436	D-003109-08	GGAACAACCTGCAGTTTGG	628
CSK						
CSK	NM_004383	4758077	1445	D-003110-05	CAGAATGTATTGCCAAGTA	629
CSK	NM_004383	4758077	1445	D-003110-06	GAACAAAGTCGCCGTCAAG	630
CSK	NM_004383	4758077	1445	D-003110-07	GCGAGTGCCTTATCCAAGA	631
CSK	NM_004383	4758077	1445	D-003110-08	GGAGAAGGGCTACAAGATG	632
DDR1						
DDR1	NM_013994	7669484	780	D-003111-05	GGAGATGGAGTTTGAGTTT	633
DDR1	NM_013994	7669484	780	D-003111-06	CAGAGGCCCTGTCATCTTT	634
DDR1	NM_013994	7669484	780	D-003111-07	GCTGGTAGCTGTCAAGATC	635
DDR1	NM_013994	7669484	780	D-003111-08	TGAAAGAGGTGAAGATCAT	636
DDR2						
DDR2	NM_006182	5453813	4921	D-003112-05	GGTAAGAACTACACAATCA	637
DDR2	NM_006182	5453813	4921	D-003112-06	GAACGAGAGTGCCACCAAT	638
DDR2	NM_006182	5453813	4921	D-003112-07	ACACCAATCTGAAGTTTAT	639
DDR2	NM_006182	5453813	4921	D-003112-08	CAACAAGAATGCCAGGAAT	640
DKFZp761P1010						
DKFZp761P1010	NM_018423	8922178	55359	D-003113-05	CCTAGAAGCTGCCATTAAA	641

DKFZp761 P1010	NM_018423	8922178	55359	D-003113-06	GATTAGGCCTGGCTTATGA	642
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-07	CCCAGTAGCTGCACACATA	643
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-08	GGTGGTACCTGAACTGTAT	644
EGFR						
EGFR	NM_005228	4885198	1956	D-003114-05	GAAGGAACTGAATTCAAA	645
EGFR	NM_005228	4885198	1956	D-003114-06	GGAAATATGTACTACGAAA	646
EGFR	NM_005228	4885198	1956	D-003114-07	CCACAAAGCAGTGAATTTA	647
EGFR	NM_005228	4885198	1956	D-003114-08	GTAACAAGCTCACGCAGTT	648
EPHA1						
EPHA1	NM_005232	4885208	2041	D-003115-05	GACCAGAGCTTCACCATTC	649
EPHA1	NM_005232	4885208	2041	D-003115-06	GCAAGACTGTGGCCATTAA	650
EPHA1	NM_005232	4885208	2041	D-003115-07	GGGCGAACCTGACCTATGA	651
EPHA1	NM_005232	4885208	2041	D-003115-08	GATTGTAGCCGTCATCTTT	652
EPHA2						
EPHA2	NM_004431	4758277	1969	D-003116-05	GGAGGGATCTGGCAACTTG	653
EPHA2	NM_004431	4758277	1969	D-003116-06	GCAGCAAGGTGCACGAATT	654
EPHA2	NM_004431	4758277	1969	D-003116-07	GGAGAAGGATGGCGAGTTC	655
EPHA2	NM_004431	4758277	1969	D-003116-08	GAAGTTCACCTACCGAGATC	656
EPHA3						
EPHA3	NM_005233	21361240	2042	D-003117-05	GATCGGACCTCCAGAAATA	657
EPHA3	NM_005233	21361240	2042	D-003117-06	GAAGTCAGCTCAGAAAGATT	658
EPHA3	NM_005233	21361240	2042	D-003117-07	GCAAGAGGCACAAATGTTA	659
EPHA3	NM_005233	21361240	2042	D-003117-08	GAGCATCAGTTTACAAAGA	660
EPHA4						
EPHA4	NM_004438	4758279	2043	D-003118-05	GGTCTGGGATGAAGTATTT	661
EPHA4	NM_004438	4758279	2043	D-003118-06	GAATGAAGTTACCTTATTG	662
EPHA4	NM_004438	4758279	2043	D-003118-07	GAAGTTGGGTGGATAGCAA	663
EPHA4	NM_004438	4758279	2043	D-003118-08	GAGATTAAATTCACCTTGA	664
EPHA7						
EPHA7	NM_004440	4758281	2045	D-003119-05	GAAAAGAGATGTTGCA GTA	665
EPHA7	NM_004440	4758281	2045	D-003119-06	CTAGATGCCTCCTGTATTA	666
EPHA7	NM_004440	4758281	2045	D-003119-07	AGAAGAAGGTTATCGTTTA	667
EPHA7	NM_004440	4758281	2045	D-003119-08	TAGCAAAGCTGACCAAGAA	668
EPHA8						
EPHA8	NM_020526	18201903	2046	D-003120-05	GAAGATGCACTATCAGAAT	669
EPHA8	NM_020526	18201903	2046	D-003120-06	GAGAAGATGCACTATCAGA	670
EPHA8	NM_020526	18201903	2046	D-003120-07	AACCTGATCTCCAGTGTGA	671
EPHA8	NM_020526	18201903	2046	D-003120-08	TCTCAGACCTGGGCTATGT	672
EPHB1						
EPHB1	NM_004441	21396502	2047	D-003121-05	GCGATAAGCTCCAGCATTAA	673
EPHB1	NM_004441	21396502	2047	D-003121-06	GAAACGGGCTTATAGCAAA	674
EPHB1	NM_004441	21396502	2047	D-003121-07	GGATGAAGATCTACATTGA	675
EPHB1	NM_004441	21396502	2047	D-003121-08	GCACGTCTCTGTCAACATC	676
EPHB2						
EPHB2	NM_017449	17975764	2048	D-003122-05	ACTATGAGCTGCAGTACTA	677
EPHB2	NM_017449	17975764	2048	D-003122-06	GTACAACGCCACAGCCATA	678
EPHB2	NM_017449	17975764	2048	D-003122-07	GGAAAGCAATGACTGTTCT	679
EPHB2	NM_017449	17975764	2048	D-003122-08	CGGACAAGCTGCAACACTA	680
EPHB3						
EPHB3	NM_004443	17975767	2049	D-003123-05	GGTGTGATCTCCAATGTGA	681
EPHB3	NM_004443	17975767	2049	D-003123-06	GGGATGACCTCCTGTACAA	682
EPHB3	NM_004443	17975767	2049	D-003123-07	CAGAAGACCTGCTCCGTAT	683
EPHB3	NM_004443	17975767	2049	D-003123-08	GAGATGAAGTACTTTGAGA	684

EPHB4						
EPHB4	NM_004444	17975769	2050	D-003124-05	GGACAAACACGGACAGTAT	685
EPHB4	NM_004444	17975769	2050	D-003124-06	GTACTAAGGTCTACATCGA	686
EPHB4	NM_004444	17975769	2050	D-003124-07	GGAGAGAAGCAGAATATTC	687
EPHB4	NM_004444	17975769	2050	D-003124-08	GCCAATAGCCACTCTAACA	688
EPHB6						
EPHB6	NM_004445	4758291	2051	D-003125-05	GGAAGTCGATCCTGCTTAT	689
EPHB6	NM_004445	4758291	2051	D-003125-06	GGACCAAGGTGGACACAAT	690
EPHB6	NM_004445	4758291	2051	D-003125-07	TGTGGGAAGTGATGAGTTA	691
EPHB6	NM_004445	4758291	2051	D-003125-08	CGGGAGACCTTCACCCTTT	692
ERBB2						
ERBB2	NM_004448	4758297	2064	D-003126-05	GGACGAATTCTGCACAATG	693
ERBB2	NM_004448	4758297	2064	D-003126-06	GACGAATTCTGCACAATGG	694
ERBB2	NM_004448	4758297	2064	D-003126-07	CTACAACACAGACACGTTT	695
ERBB2	NM_004448	4758297	2064	D-003126-08	AGACGAAGCATACGTGATG	696
ERBB3						
ERBB3	NM_001982	4503596	2065	D-003127-05	AAGAGGATGTCAACGGTTA	697
ERBB3	NM_001982	4503596	2065	D-003127-06	GAAGACTGCCAGACATTGA	698
ERBB3	NM_001982	4503596	2065	D-003127-07	GACAAACACTGGTGCTGAT	699
ERBB3	NM_001982	4503596	2065	D-003127-08	GCAGTGGATTTCGAGAAGTG	700
ERBB4						
ERBB4	NM_005235	4885214	2066	D-003128-05	GAGGAAAGATGCCAATTAA	701
ERBB4	NM_005235	4885214	2066	D-003128-06	GCAGGAAACATCTATATTA	702
ERBB4	NM_005235	4885214	2066	D-003128-07	GATCACAACCTGCTGCTTAA	703
ERBB4	NM_005235	4885214	2066	D-003128-08	CCTCAAAGATACCTAGTTA	704
FER						
FER	NM_005246	4885230	2241	D-003129-05	GGAGTGACCTGAAGAATTC	705
FER	NM_005246	4885230	2241	D-003129-06	TAAAGCAGATTCCCATTAA	706
FER	NM_005246	4885230	2241	D-003129-07	GGAAAGTACTGTCCAAATG	707
FER	NM_005246	4885230	2241	D-003129-08	GAACAACGGCTGCTAAAGA	708
FES						
FES	NM_002005	13376997	2242	D-003130-05	CGAGGATCCTGAAGCAGTA	709
FES	NM_002005	13376997	2242	D-003130-06	AGGAATACCTGGAGATTAG	710
FES	NM_002005	13376997	2242	D-003130-07	CAACAGGAGCTCCGGAATG	711
FES	NM_002005	13376997	2242	D-003130-08	GGTGTTGGGTGAGCAGATT	712
FGFR1						
FGFR1	NM_000604	13186232	2260	D-003131-05	TAAGAAATGTCTCCTTTGA	713
FGFR1	NM_000604	13186232	2260	D-003131-06	GAAGACTGCTGGAGTTAAT	714
FGFR1	NM_000604	13186232	2260	D-003131-07	GATGGTCCCTTGATGTCA	715
FGFR1	NM_000604	13186232	2260	D-003131-08	CTTAAGAAATGTCTCCTTT	716
FGFR2						
FGFR2	NM_000141	13186239	2263	D-003132-05	CCAAATCTCTCAACCAGAA	717
FGFR2	NM_000141	13186239	2263	D-003132-06	GAACAGTATTCACCTAGTT	718
FGFR2	NM_000141	13186239	2263	D-003132-07	GGCCAACACTGTCAAGTTT	719
FGFR2	NM_000141	13186239	2263	D-003132-08	GTGAAGATGTTGAAAGATG	720
FGFR3						
FGFR3	NM_000142	13112046	2261	D-003133-05	TGTCGGACCTGGTGTCTGA	721
FGFR3	NM_000142	13112046	2261	D-003133-06	GCATCAAGCTGCGGCATCA	722
FGFR3	NM_000142	13112046	2261	D-003133-07	GGACGGCACACCCTACGTT	723
FGFR3	NM_000142	13112046	2261	D-003133-08	TGCACAACCTCGACTACTA	724
FGFR4						
FGFR4	NM_002011	13112051	2264	D-003134-05	GCACTGGAGTCTCGTGATG	725
FGFR4	NM_002011	13112051	2264	D-003134-06	CATAGGGACCTCTCGAATA	726
FGFR4	NM_002011	13112051	2264	D-003134-07	ATACGGACATCATCTGTA	727
FGFR4	NM_002011	13112051	2264	D-003134-08	ATAGGGACCTCTCGAATAG	728
FGR						

FGR	NM 005248	4885234	2268	D-003135-05	GCGATCATGTGAAGCATT	729
FGR	NM 005248	4885234	2268	D-003135-06	TCACTGAGCTCATACCAA	730
FGR	NM 005248	4885234	2268	D-003135-07	GAAGAGTGGTACTTTGGAA	731
FGR	NM 005248	4885234	2268	D-003135-08	CCCAGAAGCTGCCCTCTTT	732
FLT1						
FLT1	NM 002019	4503748	2321	D-003136-05	GAGCAAACGTGACTTATTT	733
FLT1	NM 002019	4503748	2321	D-003136-06	CCAAATGGGTTTCATGTTA	734
FLT1	NM 002019	4503748	2321	D-003136-07	CAACAAGGATGCAGCACTA	735
FLT1	NM 002019	4503748	2321	D-003136-08	GGACGTAAGTGAAGAGGAT	736
FLT3						
FLT3	NM 004119	4758395	2322	D-003137-05	GAAGGCATCTACACCATT	737
FLT3	NM 004119	4758395	2322	D-003137-06	GAAGGAGTCTGGAATAGAA	738
FLT3	NM 004119	4758395	2322	D-003137-07	GAATTTAAGTCGTGTGTTT	739
FLT3	NM 004119	4758395	2322	D-003137-08	GGAATTCATTTCACTCTGA	740
FLT4						
FLT4	NM 002020	4503752	2324	D-003138-05	GCAAGAACGTGCATCTGTT	741
FLT4	NM 002020	4503752	2324	D-003138-06	GCGAATACCTGTCCTACGA	742
FLT4	NM 002020	4503752	2324	D-003138-07	GAAGACATTTGAGGAATTC	743
FLT4	NM 002020	4503752	2324	D-003138-08	GAGCAGCCATTCATCAACA	744
FRK						
FRK	NM 002031	4503786	2444	D-003139-05	GAAACAGACTCTTCATATT	745
FRK	NM 002031	4503786	2444	D-003139-06	GAACAATACCACTCCAGTA	746
FRK	NM 002031	4503786	2444	D-003139-07	CAAGACCGGTTCTTTCTA	747
FRK	NM 002031	4503786	2444	D-003139-08	GCAAGAATATCTCCAAAAT	748
FYN						
FYN	NM 002037	23510344	2534	D-003140-05	GGAATGGACTCATATGCAA	749
FYN	NM 002037	23510344	2534	D-003140-06	GCAGAAGAGTGGTACTTTG	750
FYN	NM 002037	23510344	2534	D-003140-07	CAAAGGAAGTTTACTGGAT	751
FYN	NM 002037	23510344	2534	D-003140-08	GAAGAGTGGTACTTTGGAA	752
HCK						
HCK	NM 002110	4504356	3055	D-003141-05	GAGATACCGTGAAACATTA	753
HCK	NM 002110	4504356	3055	D-003141-06	GCAGGGAGATACCGTGAAA	754
HCK	NM 002110	4504356	3055	D-003141-07	CATCGTGTTGCCCTGTAT	755
HCK	NM 002110	4504356	3055	D-003141-08	TGTGTAAGATTGCTGACTT	756
ITK						
ITK	NM 005546	21614549	3702	D-003144-05	CAAATAATCTGGAAACCTA	757
ITK	NM 005546	21614549	3702	D-003144-06	GAAGAAACGAGGAATAATA	758
ITK	NM 005546	21614549	3702	D-003144-07	GAAACTCTCTCATCCCAA	759
ITK	NM 005546	21614549	3702	D-003144-08	GGAATGGGCATGAAGGATA	760
JAK1						
JAK1	NM 002227	4504802	3716	D-003145-05	CCACATAGCTGATCTGAAA	761
JAK1	NM 002227	4504802	3716	D-003145-06	TGAAATCACTCACATTGTA	762
JAK1	NM 002227	4504802	3716	D-003145-07	TAAGGAACCTCTATCATGA	763
JAK1	NM 002227	4504802	3716	D-003145-08	GCAGGTGGCTGTAAATCT	764
JAK2						
JAK2	NM 004972	13325062	3717	D-003146-05	GCAAATAGATCCAGTTCTT	765
JAK2	NM 004972	13325062	3717	D-003146-06	GAGCAAAGATCCAAGACTA	766
JAK2	NM 004972	13325062	3717	D-003146-07	GCCAGAACTTGAACTTA	767
JAK2	NM 004972	13325062	3717	D-003146-08	GTACAGATTTGCGAGATTT	768
JAK3						
JAK3	NM 000215	4557680	3718	D-003147-05	GCGCCTATCTTTCTCCTTT	769
JAK3	NM 000215	4557680	3718	D-003147-06	CCAGAAATCGTAGACATTA	770
JAK3	NM 000215	4557680	3718	D-003147-07	CCTCATCTCTTCAGACTAT	771
JAK3	NM 000215	4557680	3718	D-003147-08	TGTACGAGCTCTTCACCTA	772
KDR						
KDR	NM 002253	11321596	3791	D-003148-05	GGAAATCTCTTGCAAGCTA	773

KDR	NM 002253	11321596	3791	D-003148-06	GATTACAGATCTCCATTTA	774
KDR	NM 002253	11321596	3791	D-003148-07	GCAGACAGATCTACGTTTG	775
KDR	NM 002253	11321596	3791	D-003148-08	GCGATGGCCTCTTCTGTAA	776
KIAA1079						
KIAA1079	NM 014916	7662475	22853	D-003149-05	GAAATTCTCTCAACTGATG	777
KIAA1079	NM 014916	7662475	22853	D-003149-06	GCAGAGGTCTTCACACTTT	778
KIAA1079	NM 014916	7662475	22853	D-003149-07	TAAATGATCTTCAGACAGA	779
KIAA1079	NM 014916	7662475	22853	D-003149-08	GAGCAGCCCTACTCTGATA	780
KIT						
KIT	NM 000222	4557694	3815	D-003150-05	AAACACGGCTTAAGCAATT	781
KIT	NM 000222	4557694	3815	D-003150-06	GAACAGAACCTTCACTGAT	782
KIT	NM 000222	4557694	3815	D-003150-07	GGGAAGCCCTCATGTCTGA	783
KIT	NM 000222	4557694	3815	D-003150-08	GCAATTCCATTTATGTGTT	784
LCK						
LCK	NM 005356	20428651	3932	D-003151-05	GAACTGCCATTATCCCATTA	785
LCK	NM 005356	20428651	3932	D-003151-06	GAGAGGTGGTGAAACATTA	786
LCK	NM 005356	20428651	3932	D-003151-07	GGGCCAAGTTTCCCATTTAA	787
LCK	NM 005356	20428651	3932	D-003151-08	GCACGCTGCTCATCCGAAA	788
LTK						
LTK	NM 002344	4505044	4058	D-003152-05	TGAATTCACCTCTGCCAAT	789
LTK	NM 002344	4505044	4058	D-003152-06	GTGGCAACCTCAACACTGA	790
LTK	NM 002344	4505044	4058	D-003152-07	GGAGCTAGCTGTGGATAAC	791
LTK	NM 002344	4505044	4058	D-003152-08	GCAAGTTTCGCCATCAGAA	792
LYN						
LYN	NM 002350	4505054	4067	D-003153-05	GCAGATGGCTTGTGCAGAA	793
LYN	NM 002350	4505054	4067	D-003153-06	GGAGAAGGCTTGTATTAGT	794
LYN	NM 002350	4505054	4067	D-003153-07	GATGAGCTCTATGACATTA	795
LYN	NM 002350	4505054	4067	D-003153-08	GGTGCTAAGTTCCTATTA	796
MATK						
MATK	NM 002378	21450841	4145	D-003154-05	TGAAGAATATCAAGTGTGA	797
MATK	NM 002378	21450841	4145	D-003154-06	CCGCTCAGCTCCTGCAGTT	798
MATK	NM 002378	21450841	4145	D-003154-07	TACTGAACCTGCAGCATTT	799
MATK	NM 002378	21450841	4145	D-003154-08	TGGGAGGTCTTCTCATATG	800
MERTK						
MERTK	NM 006343	5453737	10461	D-003155-05	GAACTTACCTTACATAGCT	801
MERTK	NM 006343	5453737	10461	D-003155-06	GGACCTGCATACTTACTTA	802
MERTK	NM 006343	5453737	10461	D-003155-07	TGACAGGAATCTTCTAATT	803
MERTK	NM 006343	5453737	10461	D-003155-08	GGTAATGGCTCAGTCATGA	804
MET						
MET	NM 000245	4557746	4233	D-003156-05	GAAAGAACCTCTCAACATT	805
MET	NM 000245	4557746	4233	D-003156-06	GGACAAGGCTGACCATATG	806
MET	NM 000245	4557746	4233	D-003156-07	CCAATGACCTGCTGAAATT	807
MET	NM 000245	4557746	4233	D-003156-08	GAGCATACATTAACCAAA	808
MST1R						
MST1R	NM 002447	4505264	4486	D-003157-05	GGATGGAGCTGCTGGCTTT	809
MST1R	NM 002447	4505264	4486	D-003157-06	CTGCAGACCTATAGATTTA	810
MST1R	NM 002447	4505264	4486	D-003157-07	GCACCTGTCTCACTCTTGA	811
MST1R	NM 002447	4505264	4486	D-003157-08	GAAAGAGTCCATCCAGCTA	812
MUSK						
MUSK	NM 005592	5031926	4593	D-003158-05	GAAGAAGCCTCGGCAGATA	813
MUSK	NM 005592	5031926	4593	D-003158-06	GTAATAATCTCCATCATGT	814
MUSK	NM 005592	5031926	4593	D-003158-07	GGAATGAACTGAAAGTAGT	815
MUSK	NM 005592	5031926	4593	D-003158-08	GAGATTTCTTGACTAGAA	816
NTRK1						
NTRK1	NM 002529	4585711	4914	D-003159-05	GGACAACCCTTTTCGAGTTC	817
NTRK1	NM 002529	4585711	4914	D-003159-06	CCAGTGACCTCAACAGGAA	818

NTRK1	NM_002529	4585711	4914	D-003159-07	CCACAATACTTCAGTGATG	819
NTRK1	NM_002529	4585711	4914	D-003159-08	GAAGAGTGGTCTCCGTTTC	820
NTRK2						
NTRK2	NM_006180	21361305	4915	D-003160-05	GAACAGAAGTAATGAAATC	821
NTRK2	NM_006180	21361305	4915	D-003160-06	GTAATGCTGTTTCTGCTTA	822
NTRK2	NM_006180	21361305	4915	D-003160-07	GCAAGACACTCCAAGTTTG	823
NTRK2	NM_006180	21361305	4915	D-003160-08	GAAAGTCTATCACATTATC	824
NTRK3						
NTRK3	NM_002530	4505474	4916	D-003161-05	GAGCGAATCTGCTAGTGAA	825
NTRK3	NM_002530	4505474	4916	D-003161-06	GAAGTTCACTACAGAGAGT	826
NTRK3	NM_002530	4505474	4916	D-003161-07	GGTCGACGGTCCAAATTTG	827
NTRK3	NM_002530	4505474	4916	D-003161-08	GAATATCACTTCCATACAC	828
PDGFRA						
PDGFRA	NM_006206	15451787	5156	D-003162-05	GAAACTTCCTGGACTATTT	829
PDGFRA	NM_006206	15451787	5156	D-003162-06	GAGATTTGGTCAACTATTT	830
PDGFRA	NM_006206	15451787	5156	D-003162-07	GCACGCCGCTTCCTGATAT	831
PDGFRA	NM_006206	15451787	5156	D-003162-08	CATCAGAGCTGGATCTAGA	832
PDGFRB						
PDGFRB	NM_002609	15451788	5159	D-003163-05	GAAAGGAGACGTCAAATAT	833
PDGFRB	NM_002609	15451788	5159	D-003163-06	GGAATGAGGTGGTCAACTT	834
PDGFRB	NM_002609	15451788	5159	D-003163-07	CAACGAGTCTCCAGTGCTA	835
PDGFRB	NM_002609	15451788	5159	D-003163-08	GAGAGGACCTGCCGAGCAA	836
PTK2						
PTK2	NM_005607	27886592	5747	D-003164-05	GAAGTTGGGTTGTCTAGAA	837
PTK2	NM_005607	27886592	5747	D-003164-06	GAAGAACAATGATGTAATC	838
PTK2	NM_005607	27886592	5747	D-003164-07	GGAAATTGCTTTGAAGTTG	839
PTK2	NM_005607	27886592	5747	D-003164-08	GGTTCAAGCTGGATTATTT	840
PTK2B						
PTK2B	NM_004103	27886583	2185	D-003165-05	GAACATGGCTGACCTCATA	841
PTK2B	NM_004103	27886583	2185	D-003165-06	GGACCACGCTGCTCTATTT	842
PTK2B	NM_004103	27886583	2185	D-003165-07	GGACGAGGACTATTACAAA	843
PTK2B	NM_004103	27886583	2185	D-003165-08	TGGCAGAGCTCATCAACAA	844
PTK6						
PTK6	NM_005975	27886594	5753	D-003166-05	GAGAAAGTCCTGCCCGTTT	845
PTK6	NM_005975	27886594	5753	D-003166-06	TGAAGAAGCTGCGGCACAA	846
PTK6	NM_005975	27886594	5753	D-003166-07	CCGCGACTCTGATGAGAAA	847
PTK6	NM_005975	27886594	5753	D-003166-08	TGCCCGAGCTTGTGAACATA	848
PTK7						
PTK7	NM_002821	27886610	5754	D-003167-05	GAGAGAAGCCCACTATTAA	849
PTK7	NM_002821	27886610	5754	D-003167-06	CGAGAGAAGCCCACTATTA	850
PTK7	NM_002821	27886610	5754	D-003167-07	GGAGGGAGTTGGAGATGTT	851
PTK7	NM_002821	27886610	5754	D-003167-08	GAAGACATGCCGCTATTTG	852
PTK9						
PTK9	NM_002822	4506274	5756	D-003168-05	GAAGAACTACGACAGATTA	853
PTK9	NM_002822	4506274	5756	D-003168-09	GAAGGAGACTATTTAGAGT	854
PTK9	NM_002822	4506274	5756	D-003168-10	GAGCGGATGCTGTATTCTA	855
PTK9	NM_002822	4506274	5756	D-003168-11	CTGCAGACTTCCTTTATGA	856
PTK9L						
PTK9L	NM_007284	31543446	11344	D-003169-05	AGAGAGAGCTCCAGCAGAT	857
PTK9L	NM_007284	31543446	11344	D-003169-06	TTAACGAGGTGAAGACAGA	858
PTK9L	NM_007284	31543446	11344	D-003169-07	ACACAGAGCCACCGGATGT	859
PTK9L	NM_007284	31543446	11344	D-003169-08	GCTGGGATCAGGACTATGA	860
RET						
RET	NM_000323	21536316	5979	D-003170-05	GCAAAGACCTGGAGAAGAT	861
RET	NM_000323	21536316	5979	D-003170-06	GCACACGGCTGCATGAGAA	862
RET	NM_000323	21536316	5979	D-003170-07	GAACTGGCCTGGAGAGAGT	863

RET	NM_000323	21536316	5979	D-003170-08	TTAAATGGATGGCAATTGA	864
ROR1						
ROR1	NM_005012	4826867	4919	D-003171-05	GCAAGCATCTTTACTAGGA	865
ROR1	NM_005012	4826867	4919	D-003171-06	GAGCAAGGCTAAAGAGCTA	866
ROR1	NM_005012	4826867	4919	D-003171-07	GAGAGCAACTTCATGTAAA	867
ROR1	NM_005012	4826867	4919	D-003171-08	GAGAATGTCCTGTGTCAA	868
ROR2						
ROR2	NM_004560	19743897	4920	D-003172-05	GGAACCTCGCTGCTGCCTAT	869
ROR2	NM_004560	19743897	4920	D-003172-06	GCAGGTGCCTCCTCAGATG	870
ROR2	NM_004560	19743897	4920	D-003172-07	GCAATGTGCTAGTGTACGA	871
ROR2	NM_004560	19743897	4920	D-003172-08	GAAGACAGAATATGGTTCA	872
ROS1						
ROS1	NM_002944	19924164	6098	D-003173-05	GAGGAGACCTTCTTACTTA	873
ROS1	NM_002944	19924164	6098	D-003173-06	TTACAGAGGTTCCAGGATTA	874
ROS1	NM_002944	19924164	6098	D-003173-07	GAACAAACCTAAGCATGAA	875
ROS1	NM_002944	19924164	6098	D-003173-08	GAAAGAGCACTTCAAATAA	876
RYK						
RYK	NM_002958	11863158	6259	D-003174-05	GAAAGATGGTTACCGAATA	877
RYK	NM_002958	11863158	6259	D-003174-06	CAAAGTAGATTCTGAAGTT	878
RYK	NM_002958	11863158	6259	D-003174-07	TCACTACGCTCTATCCTTT	879
RYK	NM_002958	11863158	6259	D-003174-08	GGTGAAGGATATAGCAATA	880
SRC						
SRC	NM_005417	21361210	6714	D-003175-05	GAGAACCTGGTGTGCAAAG	881
SRC	NM_005417	21361210	6714	D-003175-09	GAGAGAACCTGGTGTGCAA	882
SRC	NM_005417	21361210	6714	D-003175-10	GGAGTTTGCTGGACTTTCT	883
SRC	NM_005417	21361210	6714	D-003175-11	GAAAGTGAGACCACGAAAG	884
SYK						
SYK	NM_003177	21361552	6850	D-003176-05	GGAATAATCTCAAGAATCA	885
SYK	NM_003177	21361552	6850	D-003176-06	GAAGTGGGCTCTGGTAATT	886
SYK	NM_003177	21361552	6850	D-003176-07	GGAAGAATCTGAGCAAATT	887
SYK	NM_003177	21361552	6850	D-003176-08	GAACAGACATGTCAAGGAT	888
TEC						
TEC	NM_003215	4507428	7006	D-003177-05	GAAATTGTCTAGTAAGTGA	889
TEC	NM_003215	4507428	7006	D-003177-06	CACCTGAAGTGTTTAATTA	890
TEC	NM_003215	4507428	7006	D-003177-07	GTACAAAGTCGCAATCAAA	891
TEC	NM_003215	4507428	7006	D-003177-08	TGGAGGAGATTCTTATTAA	892
TEK						
TEK	NM_000459	4557868	7010	D-003178-05	GAAAGAATATGCCTCCAAA	893
TEK	NM_000459	4557868	7010	D-003178-06	GGAATGACATCAAATTTCA	894
TEK	NM_000459	4557868	7010	D-003178-07	TGAAGTACCTGATATTCTA	895
TEK	NM_000459	4557868	7010	D-003178-08	CGAAAGACCTACGTGAATA	896
TIE						
TIE	NM_005424	4885630	7075	D-003179-05	GAGAGGAGGTTTATGTGAA	897
TIE	NM_005424	4885630	7075	D-003179-06	GGGACAGCCTCTACCCTTA	898
TIE	NM_005424	4885630	7075	D-003179-07	GAAGTTCTGTGCAAATTGG	899
TIE	NM_005424	4885630	7075	D-003179-08	CAACATGGCCTCAGAACTG	900
TNK1						
TNK1	NM_003985	4507610	8711	D-003180-05	GTTCTGGGCCTAAGTCTAA	901
TNK1	NM_003985	4507610	8711	D-003180-06	GAAGTGGGCTCTACAAGATC	902
TNK1	NM_003985	4507610	8711	D-003180-07	CGAGAGGTATCGGTCATGA	903
TNK1	NM_003985	4507610	8711	D-003180-08	GGCGCATCCTGGAGCATT	904
TXK						
TXK	NM_003328	4507742	7294	D-003181-05	GAACATCTATTGAGACAAG	905
TXK	NM_003328	4507742	7294	D-003181-06	TCAAGGCACTTTATGATTT	906
TXK	NM_003328	4507742	7294	D-003181-07	GGAGAGGAATGGCTATATT	907
TXK	NM_003328	4507742	7294	D-003181-08	GGATATATGTGAAGGAATG	908

TYK2						
TYK2	NM_003331	4507748	7297	D-003182-05	GAGGAGATCCACCACTTTA	909
TYK2	NM_003331	4507748	7297	D-003182-06	GCATCCACATTGCACATAA	910
TYK2	NM_003331	4507748	7297	D-003182-07	TCAAATACCTAGCCACACT	911
TYK2	NM_003331	4507748	7297	D-003182-08	CAATCTTGCTGACGTCTTG	912
TYRO3						
TYRO3	NM_006293	27597077	7301	D-003183-05	GGTAGAAGGTGTGCCATTT	913
TYRO3	NM_006293	27597077	7301	D-003183-06	ACGCTGAGATTTACAATA	914
TYRO3	NM_006293	27597077	7301	D-003183-07	GGATGGCTCCTTTGTGAAA	915
TYRO3	NM_006293	27597077	7301	D-003183-08	GAGAGGAACCTACGAAGATC	916
YES1						
YES1	NM_005433	21071041	7525	D-003184-05	GAAGGACCCTGATGAAAGA	917
YES1	NM_005433	21071041	7525	D-003184-06	TAAGAAGGGTGAAAGATTT	918
YES1	NM_005433	21071041	7525	D-003184-07	TCAAGAAGCTCAGATAATG	919
YES1	NM_005433	21071041	7525	D-003184-08	CAGAATCCCTCCATGAATT	920

Table VIII

5

Gene Name	Acc#	GI	Locus Link	Duplex #	Full Sequence	SEQ. ID NO.
APC2						
APC2	NM_013366	7549800	29882	D-003200-05	GCAAGGACCTCTTCATCAA	921
APC2	NM_013366	7549800	29882	D-003200-06	GAGAAGAAGTCCACACTAT	922
APC2	NM_013366	7549800	29882	D-003200-07	GGAATGCCATCTCCCAATG	923
APC2	NM_013366	7549800	29882	D-003200-09	CAACACGTGTGACATCATC	924
ATM						
ATM	NM_000051	20336202	472	D-003201-05	GCAAGCAGCTGAAACAAAT	925
ATM	NM_000051	20336202	472	D-003201-06	GAATGTTGCTTTCTGAATT	926
ATM	NM_000051	20336202	472	D-003201-07	GACCTGAAGTCTTATTTAA	927
ATM	NM_000051	20336202	472	D-003201-08	AGACAGAATTCCCAAATAA	928
ATR						
ATR	NM_001184	20143978	545	D-003202-05	GAACAACACTGCTGGTTTG	929
ATR	NM_001184	20143978	545	D-003202-06	GAAGTCATCTGTTCAATTAT	930
ATR	NM_001184	20143978	545	D-003202-07	GAAATAAGGTAGACTCAAT	931
ATR	NM_001184	20143978	545	D-003202-08	CAACATAAATCCAAGAAGA	932
BTAK						
BTAK	NM_003600	3213196	6790	D-003545-04	CAAAGAATCAGCTAGCAAA	933
BTAK	NM_003600	3213196	6790	D-003203-05	GAAGAGAGTTATTCATAGA	934
BTAK	NM_003600	3213196	6790	D-003203-07	CAAATGCCCTGTCTTACTG	935
STK6	NM_003600	3213196	6790	D-003203-09	TCTCGTGACTCAGCAAATT	936
CCNA1						
CCNA1	NM_003914	16306528	8900	D-003204-05	GAACCTGGCTAAGTACGTA	937
CCNA1	NM_003914	16306528	8900	D-003204-06	GCAGATCCATTCTTGAAAT	938
CCNA1	NM_003914	16306528	8900	D-003204-07	TCACAAGAATCAGGTGTTA	939
CCNA1	NM_003914	16306528	8900	D-003204-08	CATAAAGCGTACCTTGATA	940
CCNA2						
CCNA2	NM_001237	16950653	890	D-003205-05	GCTGTGAACTACATTGATA	941
CCNA2	NM_001237	16950653	890	D-003205-06	GATGATACCTACACCAAGA	942
CCNA2	NM_001237	16950653	890	D-003205-07	GCTGTTAGCCTCAAAGTTT	943
CCNA2	NM_001237	16950653	890	D-003205-08	AAGCTGGCCTGAATCATT	944
CCNB1						
CCNB1	NM_031966	14327895	891	D-003206-05	CAACATTACCTGTCATATA	945
CCNB1	NM_031966	14327895	891	D-003206-06	CCAAATACCTGATGGAAC	946
CCNB1	NM_031966	14327895	891	D-003206-07	GAAATGTACCCTCCAGAAA	947

CCNB1	NM_031966	14327895	891	D-003206-08	GCACCTGGCTAAGAATGTA	948
CCNB2						
CCNB2	NM_004701	10938017	9133	D-003207-05	CAACAAATGTCAACAAACA	949
CCNB2	NM_004701	10938017	9133	D-003207-06	GCAGCAAACCTCCTGAAGAT	950
CCNB2	NM_004701	10938017	9133	D-003207-07	CCAGTGATTTGGAGAATAT	951
CCNB2	NM_004701	10938017	9133	D-003207-08	GTGACTACGTAAAGGATAT	952
CCNB3						
CCNB3	NM_033031	14719419	85417	D-003208-05	TGAACAAACTGCTGACTTT	953
CCNB3	NM_033031	14719419	85417	D-003208-06	GCTAGCTGCTGCCTCCTTA	954
CCNB3	NM_033031	14719419	85417	D-003208-07	CAACTCACCTCGTGTGGAT	955
CCNB3	NM_033031	14719419	85417	D-003208-08	GTGGATCTCTACCTAATGA	956
CCNC						
CCNC	NM_005190	7382485	892	D-003209-05	GCAGAGCTCCCACTATTTG	957
CCNC	NM_005190	7382485	892	D-003209-06	GGAGTAGTTTCAAATACAA	958
CCNC	NM_005190	7382485	892	D-003209-07	GACCTTTGCTCCAGTATGT	959
CCNC	NM_005190	7382485	892	D-003209-08	GAGATTCTATGCCAGGTAT	960
CCND1						
CCND1	NM_053056	16950654	595	D-003210-05	TGAACAAGCTCAAGTGGAA	961
CCND1	NM_053056	16950654	595	D-003210-06	CCAGAGTGATCAAGTGTGA	962
CCND1	NM_053056	16950654	595	D-003210-07	GTTCTGTGGCCTCTAAGATG	963
CCND1	NM_053056	16950654	595	D-003210-08	CCGAGAAGCTGTGCATCTA	964
CCND2						
CCND2	NM_001759	16950656	894	D-003211-06	TGAATTACCTGGACCGTTT	965
CCND2	NM_001759	16950656	894	D-003211-07	CGGAGAAGCTGTGCATTTA	966
CCND2	NM_001759	16950656	894	D-003211-08	CTACAGACGTGCGGGATAT	967
CCND2	NM_001759	16950656	894	D-003211-09	CAACACAGACGTGGATTGT	968
CCND3						
CCND3	NM_001760	16950657	896	D-003212-05	GGACCTGGCTGCTGTGATT	969
CCND3	NM_001760	16950657	896	D-003212-06	GATTATACCTTTGCCATGT	970
CCND3	NM_001760	16950657	896	D-003212-07	GACCAGCACTCCTACAGAT	971
CCND3	NM_001760	16950657	896	D-003212-08	TGCGGAAGATGCTGGCTTA	972
CCNE1						
CCNE1	NM_001238	17318558	898	D-003213-05	GTACTGAGCTGGGCAAATA	973
CCNE1	NM_001238	17318558	898	D-003213-06	GGAAATCTATCCTCCAAAG	974
CCNE1	NM_001238	17318558	898	D-003213-07	GGAGGTGTGTGAAGTCTAT	975
CCNE1	NM_001238	17318558	898	D-003213-08	CTAAATGACTTACATGAAG	976
CCNE2						
CCNE2	NM_057749	17318564	9134	D-003214-05	GGATGGAACTCATTATATT	977
CCNE2	NM_057749	17318564	9134	D-003214-06	GCAGATATGTTTCATGACAA	978
CCNE2	NM_057749	17318564	9134	D-003214-07	CATAATATCCAGACACATA	979
CCNE2	NM_057749	17318564	9134	D-003214-08	TAAGAAAGCCTCAGGTTTG	980
CCNF						
CCNF	NM_001761	4502620	899	D-003215-05	TCACAAAGCATCCATATTG	981
CCNF	NM_001761	4502620	899	D-003215-06	GAAGTCATGTTTACAGTGT	982
CCNF	NM_001761	4502620	899	D-003215-07	TAGCCTACCTCTACAATGA	983
CCNF	NM_001761	4502620	899	D-003215-08	GCACCCGGTTTATCAGTAA	984
CCNG1						
CCNG1	NM_004060	8670528	900	D-003216-05	GATAATGGCCTCAGAATGA	985
CCNG1	NM_004060	8670528	900	D-003216-06	GCACGGCAATTGAAGCATA	986
CCNG1	NM_004060	8670528	900	D-003216-07	GGAATAGAATGTCTTCAGA	987
CCNG1	NM_004060	8670528	900	D-003216-08	TAACTCACCTTCCAACAAT	988
CCNG2						
CCNG2	NM_004354	4757935	901	D-003217-05	GGAGAGAGTTGGTTTCTAA	989
CCNG2	NM_004354	4757935	901	D-003217-06	GGTGAAACCTAAACATTG	990
CCNG2	NM_004354	4757935	901	D-003217-07	GAAATACTGAGCCTTGATA	991
CCNG2	NM_004354	4757935	901	D-003217-08	TGCCAAAGTTGAAGATTTA	992

CCNH						
CCNH	NM 001239	17738313	902	D-003218-05	GCTGATGACTTTCTTAATA	993
CCNH	NM 001239	17738313	902	D-003218-06	CAACTTAATTTCCACCTTA	994
CCNH	NM 001239	17738313	902	D-003218-07	ATACACACCTTCCCAAATT	995
CCNH	NM 001239	17738313	902	D-003218-08	GCTATGAAGATGATGATTA	996
CCNI						
CCNI	NM 006835	17738314	10983	D-003219-05	GCAAGCAGACCTCTACTAA	997
CCNI	NM 006835	17738314	10983	D-003219-07	TGAGAGAATTCCAGTACTA	998
CCNI	NM 006835	17738314	10983	D-003219-08	GGAATCAAACGGCTCTATA	999
CCNI	NM 006835	17738314	10983	D-003219-09	GAATTGGGATCTTCACACA	1000
CCNT1						
CCNT1	NM 001240	17978465	904	D-003220-05	TATCAACACTGCTATAGTA	1001
CCNT1	NM 001240	17978465	904	D-003220-06	GAACAAACGTCCTGGTGAT	1002
CCNT1	NM 001240	17978465	904	D-003220-07	GCACAAGACTCACCCATCT	1003
CCNT1	NM 001240	17978465	904	D-003220-08	GCACAGACTTCTTACTTCA	1004
CCNT2A						
CCNT2A	NM 001241	17978467	905	D-003221-05	GCACAGACATCCTATTTCA	1005
CCNT2A	NM 001241	17978467	905	D-003221-06	GCAGGGACCTTCTATATCA	1006
CCNT2A	NM 001241	17978467	905	D-003221-07	GAACAGCTATATTCACAGA	1007
CCNT2A	NM 001241	17978467	905	D-003221-09	TTATATAGCTGCCCAGGTA	1008
CCNT2B						
CCNT2B	NM 058241	17978468	905	D-003222-05	GCACAGACATCCTATTTCA	1009
CCNT2B	NM 058241	17978468	905	D-003222-06	GCAGGGACCTTCTATATCA	1010
CCNT2B	NM 058241	17978468	905	D-003222-07	GAACAGCTATATTCACAGA	1011
CCNT2B	NM 058241	17978468	905	D-003222-08	GGTGAAATGTACCCAGTTA	1012
CDC16						
CDC16	NM 003903	14110370	8881	D-003223-05	GTAGATGGCTTGCAAGAGA	1013
CDC16	NM 003903	14110370	8881	D-003223-06	TAAAGTAGCTTCACTCTCT	1014
CDC16	NM 003903	14110370	8881	D-003223-07	GCTACAAGCTTACTTCTGT	1015
CDC16	NM 003903	14110370	8881	D-003223-08	TGGAAGAGCCCATCAATAA	1016
CDC2						
CDC2	NM 033379	27886643	983	D-003552-01	GTACAGATCTCCAGAAGTA	1017
CDC2	NM 033379	27886643	983	D-003552-02	GATCAACTCTTCAGGATTT	1018
CDC2	NM 033379	27886643	983	D-003552-03	GGTTATATCTCATCTTTGA	1019
CDC2	NM 033379	27886643	983	D-003552-04	GAACCTTCGTCATCCAAATA	1020
CDC20						
CDC20	NM 001255	4557436	991	D-003225-05	GGGAATATATATCCTCTGT	1021
CDC20	NM 001255	4557436	991	D-003225-06	GAAACGGCTTCGAAATATG	1022
CDC20	NM 001255	4557436	991	D-003225-07	GAAGACCTGCCGTTACATT	1023
CDC20	NM 001255	4557436	991	D-003225-08	CACCAGTGATCGACACATT	1024
CDC25A						
CDC25A	NM 001789	4502704	993	D-003226-05	GAAATTATGGCATCTGTTT	1025
CDC25A	NM 001789	4502704	993	D-003226-06	TACAAGGAGTTCTTTATGA	1026
CDC25A	NM 001789	4502704	993	D-003226-07	CCACGAGGACTTTAAAGAA	1027
CDC25A	NM 001789	4502704	993	D-003226-08	TGGGAAACATCAGGATTTA	1028
CDC25B						
CDC25B	NM 004358	11641416	994	D-003227-05	GCAGATACCCCTATGAATA	1029
CDC25B	NM 004358	11641416	994	D-003227-06	CTAGGTCGCTTCTCTCTGA	1030
CDC25B	NM 004358	11641416	994	D-003227-07	GAGAGCTGATTGGAGATTA	1031
CDC25B	NM 004358	11641416	994	D-003227-08	AAAAGGACCTCGTCATGTA	1032
CDC25C						
CDC25C	NM 001790	12408659	995	D-003228-05	GAGCAGAAGTGGCCTATAT	1033
CDC25C	NM 001790	12408659	995	D-003228-06	CAGAAGAGATTTAGATGA	1034
CDC25C	NM 001790	12408659	995	D-003228-07	CCAGGGAGCCTTAACTTA	1035
CDC25C	NM 001790	12408659	995	D-003228-08	GAAACTTGGTGGACAGTGA	1036
CDC27						

CDC27	NM 001256	16554576	996	D-003229-06	CATGCAAGCTGAAAGAATA	1037
CDC27	NM 001256	16554576	996	D-003229-07	CAACACAAGTACCTAATCA	1038
CDC27	NM 001256	16554576	996	D-003229-08	GGAGATGGATCCTATTTAC	1039
CDC27	NM 001256	16554576	996	D-003229-09	GAAAAGCCATGATGATATT	1040
CDC34						
CDC34	NM 004359	16357476	997	D-003230-05	GCTCAGACCTCTTCTACGA	1041
CDC34	NM 004359	16357476	997	D-003230-06	GGACGAGGGCGATCTATAC	1042
CDC34	NM 004359	16357476	997	D-003230-07	GATCGGGAGTACACAGACA	1043
CDC34	NM 004359	16357476	997	D-003230-08	TGAACGAGCCCAACACCTT	1044
CDC37						
CDC37	NM 007065	16357478	11140	D-003231-05	GCGAGGAGACAGCCAATTA	1045
CDC37	NM 007065	16357478	11140	D-003231-06	CACAAGACCTTCGTGGAAA	1046
CDC37	NM 007065	16357478	11140	D-003231-07	ACAATCGTCATGCAATTTA	1047
CDC37	NM 007065	16357478	11140	D-003231-08	GAGGAGAAATGTGCACTCA	1048
CDC45L						
CDC45L	NM 003504	34335230	8318	D-003232-05	GCACACGGATCTCCTTTGA	1049
CDC45L	NM 003504	34335230	8318	D-003232-06	GCAAACACCTGCTCAAGTC	1050
CDC45L	NM 003504	34335230	8318	D-003232-07	TGAAGAGTCTGCAAATAAA	1051
CDC45L	NM 003504	34335230	8318	D-003232-08	GGACGTGGATGCTCTGTGT	1052
CDC6						
CDC6	NM 001254	16357469	990	D-003233-05	GAACACAGCTGTCCCAGAT	1053
CDC6	NM 001254	16357469	990	D-003233-06	GAGCAGAGATGTCCACTGA	1054
CDC6	NM 001254	16357469	990	D-003233-07	GGAAATATCTTAGCTACTG	1055
CDC6	NM 001254	16357469	990	D-003233-08	GGACGAAGATTGGTATTTG	1056
CDC7						
CDC7	NM 003503	11038647	8317	D-003234-05	GGAATGAGGTACCTGATGA	1057
CDC7	NM 003503	11038647	8317	D-003234-06	CAGGAAAGGTGTTACAAA	1058
CDC7	NM 003503	11038647	8317	D-003234-07	CTACACAAATGCACAAATT	1059
CDC7	NM 003503	11038647	8317	D-003234-08	GTACGGGAATATATGCTTA	1060
CDK10						
CDK10	NM 003674	32528262	8558	D-003235-05	GAAGTGTGTTGGGAACCA	1061
CDK10	NM 003674	32528262	8558	D-003235-06	GGAAGCAGCCCTACAACAA	1062
CDK10	NM 003674	32528262	8558	D-003235-07	GCACGCCAGTGAGAACAT	1063
CDK10	NM 003674	32528262	8558	D-003235-08	GGAAGCAGCCCTACAACAA	1064
CDK2						
CDK2	NM 001798	16936527	1017	D-003236-05	GAGCTTAACCATCCTAATA	1065
CDK2	NM 001798	16936527	1017	D-003236-06	GAGCTTAACCATCCTAATA	1066
CDK2	NM 001798	16936527	1017	D-003236-07	GTACCGAGCTCCTGAAATC	1067
CDK2	NM 001798	16936527	1017	D-003236-08	GAGAGGTGGTGGCGCTTA	1068
CDK3						
CDK3	NM 001258	4557438	1018	D-003237-05	GAGCATTGGTTGCATCTTT	1069
CDK3	NM 001258	4557438	1018	D-003237-06	GATCGGAGAGGGCACCTAT	1070
CDK3	NM 001258	4557438	1018	D-003237-07	GAAGCTCTATCTGGTGT	1071
CDK3	NM 001258	4557438	1018	D-003237-08	GCAGAGATGGTGAACGAA	1072
CDK4						
CDK4	NM 000075	456426	1019	D-003238-05	GCAGCACTCTTATCTACAT	1073
CDK4	NM 000075	456426	1019	D-003238-06	GGAGGAGGCCTTCCCATCA	1074
CDK4	NM 000075	456426	1019	D-003238-07	TCGAAAGCCTCTCTTCTGT	1075
CDK4	NM 000075	456426	1019	D-003238-08	GTACCGAGCTCCCGAAGTT	1076
CDK5						
CDK5	NM 004935	4826674	1020	D-003239-05	TGACCAAGCTGCCAGACTA	1077
CDK5	NM 004935	4826674	1020	D-003239-06	GAGCTGAAATTGGCTGATT	1078
CDK5	NM 004935	4826674	1020	D-003239-07	CAACATCCCTGGTGAACGT	1079
CDK5	NM 004935	4826674	1020	D-003239-08	GGATTCCCGTCCGCTGTTA	1080
CDK6						
CDK6	NM 001259	16950658	1021	D-003240-05	GCAAAGACCTACTTCTGAA	1081

CDK6	NM 001259	16950658	1021	D-003240-06	GAAGAAGACTGGCCTAGAG	1082
CDK6	NM 001259	16950658	1021	D-003240-07	GGTCTGGACTTTCTTCATT	1083
CDK6	NM 001259	16950658	1021	D-003240-08	TAACAGATATCGATGAACT	1084
CDK7						
CDK7	NM 001799	16950659	1022	D-003241-05	GGACATAGATCAGAAGCTA	1085
CDK7	NM 001799	16950659	1022	D-003241-06	CAATAGAGCTTATACACAT	1086
CDK7	NM 001799	16950659	1022	D-003241-07	CATACAAGGCTTATTCTTA	1087
CDK7	NM 001799	16950659	1022	D-003241-08	GGAGACGACTTACTAGATC	1088
CDK8						
CDK8	NM 001260	4502744	1024	D-003242-05	CCACAGTACTCACATCAGA	1089
CDK8	NM 001260	4502744	1024	D-003242-06	GCAATAACCACACTAATGG	1090
CDK8	NM 001260	4502744	1024	D-003242-07	GAAGAAAGTGAGAGTTGTT	1091
CDK8	NM 001260	4502744	1024	D-003242-08	GAACATGACCTCTGGCATA	1092
CDK9						
CDK9	NM 001261	17017983	1025	D-003243-05	GGCCAAACGTGGACAACATA	1093
CDK9	NM 001261	17017983	1025	D-003243-06	TGACGTCCATGTTGAGTA	1094
CDK9	NM 001261	17017983	1025	D-003243-07	CCAACCAGACGGAGTTTGA	1095
CDK9	NM 001261	17017983	1025	D-003243-08	GAAGGTGGCTCTGAAGAAG	1096
CDKN1C						
CDKN1C	NM 000076	4557440	1028	D-003244-05	GACCAGAACCGCTGGGATT	1097
CDKN1C	NM 000076	4557440	1028	D-003244-06	GGACCGAAGTGACAGCGA	1098
CDKN1C	NM 000076	4557440	1028	D-003244-08	GCAAGAGATCAGCGCCTGA	1099
CDKN1C	NM 000076	4557440	1028	D-003244-09	CCGCTGGGATTACGACTTC	1100
CDKN2B						
CDKN2B	NM 004936	17981693	1030	D-003245-05	GCGAGGAGAACAAGGGCAT	1101
CDKN2B	NM 004936	17981693	1030	D-003245-06	CCAACGGAGTCAACCGTTT	1102
CDKN2B	NM 004936	17981693	1030	D-003245-07	CGATCCAGGTCATGATGAT	1103
CDKN2B	NM 004936	17981693	1030	D-003245-08	CCTGGAAGCCGGCGCGGAT	1104
CDKN2C						
CDKN2C	NM 001262	17981697	1031	D-003246-05	GGACACCGCCTGTGATTTG	1105
CDKN2C	NM 001262	17981697	1031	D-003246-06	GCCAGGAGACTGCTACTTA	1106
CDKN2C	NM 001262	17981697	1031	D-003246-07	TGAAAGACCGAACTGGTTT	1107
CDKN2C	NM 001262	17981697	1031	D-003246-08	GAACCTGCCCTTGCACTTG	1108
CDKN2D						
CDKN2D	NM 001800	17981700	1032	D-003247-05	TGGCAGTTCAAGAGGGTCA	1109
CDKN2D	NM 001800	17981700	1032	D-003247-06	CTCAGGACCTCGTGGACAT	1110
CDKN2D	NM 001800	17981700	1032	D-003247-07	TGAAGGTCCTAGTGGAGCA	1111
CDKN2D	NM 001800	17981700	1032	D-003247-08	AGACGGCGCTGCAGGTCAT	1112
CDT1						
CDT1	NM 030928	19923847	81620	D-003248-05	CCAAGGAGGCACAGAAGCA	1113
CDT1	NM 030928	19923847	81620	D-003248-06	GCTTCAACGTGGATGAAGT	1114
CDT1	NM 030928	19923847	81620	D-003248-07	TCTCCGGGCCAGAAGATAA	1115
CDT1	NM 030928	19923847	81620	D-003248-08	GCGCAATGTTGGCCAGATC	1116
CENPA						
CENPA	NM 001809	4585861	1058	D-003249-05	GCACACACCTCTTGATAAG	1117
CENPA	NM 001809	4585861	1058	D-003249-06	GCAAGAGAAATATGTGTTA	1118
CENPA	NM 001809	4585861	1058	D-003249-07	TTACATGCAGGCCGAGTTA	1119
CENPA	NM 001809	4585861	1058	D-003249-08	GAGACAAGGTTGGCTAAAG	1120
CENPB						
CENPB	NM 001810	26105977	1059	D-003250-05	GGACATAGCCGCCTGCTTT	1121
CENPB	NM 001810	26105977	1059	D-003250-06	GCACGATCCTGAAGAACAA	1122
CENPB	NM 001810	26105977	1059	D-003250-07	GGAGGAGGGTGATGTTGAT	1123
CENPB	NM 001810	26105977	1059	D-003250-08	CCGAATGGCTGCAGAGTCT	1124
CENPC1						
CENPC1	NM 001812	4502778	1060	D-003251-05	GCGAATAGATTATCAAGGA	1125
CENPC1	NM 001812	4502778	1060	D-003251-06	GAACAGAATCCATCACAAA	1126

CENPC1	NM 001812	4502778	1060	D-003251-07	CCATAAACCTCACCCAGTA	1127
CENPC1	NM 001812	4502778	1060	D-003251-08	CAAGAGAACACGTTTGAAA	1128
CENPE						
CENPE	NM 001813	4502780	1062	D-003252-05	GAAGACAGCTCAAATAATA	1129
CENPE	NM 001813	4502780	1062	D-003252-06	CAACAAAGCTACTAAATCA	1130
CENPE	NM 001813	4502780	1062	D-003252-07	GGAAAGAAAGTGCTACCATA	1131
CENPE	NM 001813	4502780	1062	D-003252-08	GGAAAGAAATGACACAGTT	1132
CENPF						
CENPF	NM 016343	14670380	1063	D-003253-05	GCGAATATCTGAATTAGAA	1133
CENPF	NM 016343	14670380	1063	D-003253-06	GGAAATTAATGCATCCTTA	1134
CENPF	NM 016343	14670380	1063	D-003253-07	GAGCGAGGCTGGTGGTTTA	1135
CENPF	NM 016343	14670380	1063	D-003253-08	CAAGTCATCTTTCATCTAA	1136
CENPH						
CENPH	NM 022909	21264590	64946	D-003254-05	GAAAGAAGAGATTGCAATT	1137
CENPH	NM 022909	21264590	64946	D-003254-06	CAGAACAAATTATGCAAGA	1138
CENPH	NM 022909	21264590	64946	D-003254-07	CTAGTGTGCTCATGGATAA	1139
CENPH	NM 022909	21264590	64946	D-003254-08	GAAACACCTATTAGAGCTA	1140
CHEK1						
CHEK1	NM 001274	20127419	1111	D-003255-05	CAAATTGGATGCAGACAAA	1141
CHEK1	NM 001274	20127419	1111	D-003255-06	GCAACAGTATTTCCGTATA	1142
CHEK1	NM 001274	20127419	1111	D-003255-07	GGACTTCTCTCCAGTAAAC	1143
CHEK1	NM 001274	20127419	1111	D-003255-08	AAAGATAGATGGTACAACA	1144
CHEK2						
CHEK2	NM 007194	22209010	11200	D-003256-02	CTCTTACATTGCATACATA	1145
CHEK2	NM 007194	22209010	11200	D-003256-03	TAAACGCCTGAAAGAAGCT	1146
CHEK2	NM 007194	22209010	11200	D-003256-04	GCATAGGACTCAAGTGTCA	1147
CHEK2	NM 007194	22209010	11200	D-003256-05	GAAATTGCACTGTCACTAA	1148
CNK						
CNK	NM 004073	4758015	1263	D-003257-05	GCGAGAAGATCCTAAATGA	1149
CNK	NM 004073	4758015	1263	D-003257-07	GCAAGTGGGTTGACTACTC	1150
CNK	NM 004073	4758015	1263	D-003257-08	GCACATCCGTTGGCCATCA	1151
CNK	NM 004073	4758015	1263	D-003257-09	GACCTCAAGTTGGGAAATT	1152
CRI1						
CRI1	NM 014335	7656937	23741	D-003258-05	GTGATGAGATTATTGATAG	1153
CRI1	NM 014335	7656937	23741	D-003258-06	GGACGAGGGCGAGGAATTT	1154
CRI1	NM 014335	7656937	23741	D-003258-07	GGAACGGAGCCTTGCTAA	1155
CRI1	NM 014335	7656937	23741	D-003258-08	TCAATCGTCTGACCGAAGA	1156
E2F1						
E2F1	NM 005225	12669910	1869	D-003259-05	GAACAGGGCCACTGACTCT	1157
E2F1	NM 005225	12669910	1869	D-003259-06	TGGACCACCTGATGAATAT	1158
E2F1	NM 005225	12669910	1869	D-003259-07	CCCAGGAGGTCACTTCTGA	1159
E2F1	NM 005225	12669910	1869	D-003259-08	GGCTGGACCTGGAACTGA	1160
E2F2						
E2F2	NM 004091	34485718	1870	D-003260-05	GGGAGAAGACTCGGTATGA	1161
E2F2	NM 004091	34485718	1870	D-003260-06	GAGGACAACCTGCAGATAT	1162
E2F2	NM 004091	34485718	1870	D-003260-07	TGAAGGAGCTGATGAACAC	1163
E2F2	NM 004091	34485718	1870	D-003260-08	CCAAGAAGTTCATTTACCT	1164
E2F3						
E2F3	NM 001949	12669913	1871	D-003261-05	GAAATTAGATGAACTGATC	1165
E2F3	NM 001949	12669913	1871	D-003261-06	TGAAGTGCCTGACTCAATA	1166
E2F3	NM 001949	12669913	1871	D-003261-07	GAACAAGGCAGCAGAAGTG	1167
E2F3	NM 001949	12669913	1871	D-003261-08	GAAACACACAGTCCAATGA	1168
E2F4						
E2F4	NM 001950	12669914	1874	D-003262-05	GGAGATTGCTGACAACTG	1169
E2F4	NM 001950	12669914	1874	D-003262-06	GAAGGTATCGGGCTAATCG	1170
E2F4	NM 001950	12669914	1874	D-003262-07	GTGCAGAAGTCCAGGGAAT	1171

E2F4	NM_001950	12669914	1874	D-003262-08	GGACAGTGGTGAGCTCAGT	1172
E2F5						
E2F5	NM_001951	12669916	1875	D-003263-05	GCAGATGACTACAACTTTA	1173
E2F5	NM_001951	12669916	1875	D-003263-06	GACATCAGCTACAGATATA	1174
E2F5	NM_001951	12669916	1875	D-003263-07	CAACATGTCTCTGAAAGAA	1175
E2F5	NM_001951	12669916	1875	D-003263-08	GAAGACATCTGTAATTGCT	1176
E2F6						
E2F6	NM_001952	12669917	1876	D-003264-05	TAAACAAGGTTGCAACGAA	1177
E2F6	NM_001952	12669917	1876	D-003264-06	TAGCATATGTGACCTATCA	1178
E2F6	NM_001952	12669917	1876	D-003264-07	GAAACCAGATTGGATGTTT	1179
E2F6	NM_001952	12669917	1876	D-003264-09	GGAACCTTCTGACTTATCA	1180
FOS						
FOS	NM_005252	6552332	2353	D-003265-05	GGGATAGCCTCTCTTACTA	1181
FOS	NM_005252	6552332	2353	D-003265-06	GAACAGTTATCTCCAGAAG	1182
FOS	NM_005252	6552332	2353	D-003265-07	GGAGACAGACCAACTAGAA	1183
FOS	NM_005252	6552332	2353	D-003265-08	AGACCGAGCCCTTTGATGA	1184
HIPK2						
HIPK2	NM_022740	13430859	28996	D-003266-06	GAGAATCACTCCAATCGAA	1185
HIPK2	NM_022740	13430859	28996	D-003266-07	AGACAGGGATTAAGTCAAA	1186
HIPK2	NM_022740	13430859	28996	D-003266-08	GGACAAAGACAAGTAGGTT	1187
HIPK2	NM_022740	13430859	28996	D-003266-09	GCACACACGTCAAATCATG	1188
HUS1						
HUS1	NM_004507	31077213	3364	D-003267-05	ACAAAGGCCTTATGCAATA	1189
HUS1	NM_004507	31077213	3364	D-003267-06	GAAGTGACATAGATATTA	1190
HUS1	NM_004507	31077213	3364	D-003267-07	AAGCTTAACCTCATCCTTT	1191
HUS1	NM_004507	31077213	3364	D-003267-08	GAACCTCTTCAACGAATTT	1192
JUN						
JUN	NM_002228	7710122	3725	D-003268-05	TGGAAACGACCTTCTATGA	1193
JUN	NM_002228	7710122	3725	D-003268-06	GAAGTGACAGCCAGAACA	1194
JUN	NM_002228	7710122	3725	D-003268-07	GAGCTGGAGCGCCTGATAA	1195
JUN	NM_002228	7710122	3725	D-003268-08	TAACGCAGCAGTTGCAAAC	1196
JUNB						
JUNB	NM_002229	4504808	3726	D-003269-05	GCATCAAAGTGGAGCGCAA	1197
JUNB	NM_002229	4504808	3726	D-003269-06	TGGAAGACCAAGAGCGCAT	1198
JUNB	NM_002229	4504808	3726	D-003269-07	CATACACAGCTACGGGATA	1199
JUNB	NM_002229	4504808	3726	D-003269-08	CCATCAACATGGAAGACCA	1200
LOC510 53						
LOC510 53	NM_015895	20127542	51053	D-003270-05	GGAGAAAGGCGCTGTATGA	1201
LOC510 53	NM_015895	20127542	51053	D-003270-06	GAATAGTTCTGTCCCAAGA	1202
LOC510 53	NM_015895	20127542	51053	D-003270-07	GAACATGTACAGTATATGG	1203
LOC510 53	NM_015895	20127542	51053	D-003270-08	GCAGAAACAAGAAGAAATC	1204
MAD2L1						
MAD2L1	NM_002358	6466452	4085	D-003271-05	GAAAGATGGCAGTTTGATA	1205
MAD2L1	NM_002358	6466452	4085	D-003271-06	TAAATAATGTGGTGGAACA	1206
MAD2L1	NM_002358	6466452	4085	D-003271-07	GAAATCCGTTCAAGTATCA	1207
MAD2L1	NM_002358	6466452	4085	D-003271-08	TTACTCGAGTGCAGAAATA	1208
MAD2L2						
MAD2L2	NM_006341	6006019	10459	D-003272-05	GGAAGAGCGCGCTCATAAA	1209
MAD2L2	NM_006341	6006019	10459	D-003272-06	TGGAAGAGCGCGCTCATAA	1210
MAD2L2	NM_006341	6006019	10459	D-003272-07	AGCCACTCCTGGAGAAGAA	1211
MAD2L2	NM_006341	6006019	10459	D-003272-08	TGGAGAAATTCGTCTTTGA	1212
MCM2						

MCM2	NM 004526	33356546	4171	D-003273-05	GAAGATCTTTGCCAGCATT	1213
MCM2	NM 004526	33356546	4171	D-003273-06	GGATAAGGCTCGTCAGATC	1214
MCM2	NM 004526	33356546	4171	D-003273-07	CAGAGCAGGTGACATATCA	1215
MCM2	NM 004526	33356546	4171	D-003273-08	GCCGTGGGCTCCTGTATGA	1216
MCM3						
MCM3	NM 002388	33356548	4172	D-003274-05	GGACATCAATATTCTTCTA	1217
MCM3	NM 002388	33356548	4172	D-003274-06	GCCAGGACATCTCCAGTTA	1218
MCM3	NM 002388	33356548	4172	D-003274-07	GCAGGTATGACCAGTATAA	1219
MCM3	NM 002388	33356548	4172	D-003274-08	GGAAATGCCTCAAGTACAC	1220
MCM4						
MCM4	XM 030274	22047061	4173	D-003275-05	GGACATATCTATTCTTACT	1221
MCM4	XM 030274	22047061	4173	D-003275-06	GATGTTAGTTCACCACTGA	1222
MCM4	XM 030274	22047061	4173	D-003275-07	CCAGCTGCCTCATACTTTA	1223
MCM4	XM 030274	22047061	4173	D-003275-08	GAAAGTACAAGATCGGTAT	1224
MCM5						
MCM5	NM 006739	23510447	4174	D-003276-05	GAAGATCCCTGGCATCATC	1225
MCM5	NM 006739	23510447	4174	D-003276-06	GAACAGGGTTACCATCATG	1226
MCM5	NM 006739	23510447	4174	D-003276-07	GGACAACATTGACTTCATG	1227
MCM5	NM 006739	23510447	4174	D-003276-08	CCAAGGAGGTAGCTGATGA	1228
MCM6						
MCM6	NM 005915	33469920	4175	D-003277-05	GGAAAGAGCTCAGAGATGA	1229
MCM6	NM 005915	33469920	4175	D-003277-06	GAGCAGCGATGGAGAAATT	1230
MCM6	NM 005915	33469920	4175	D-003277-07	GGAAACACCTGATGTCAAT	1231
MCM6	NM 005915	33469920	4175	D-003277-08	CCAAACATCTGCCGAAATC	1232
MCM7						
MCM7	NM 005916	33469967	4176	D-003278-05	GGAAATATCCCTCGTAGTA	1233
MCM7	NM 005916	33469967	4176	D-003278-06	GGAAGAAGCAGTTCAAGTA	1234
MCM7	NM 005916	33469967	4176	D-003278-07	CAACAAGCCTCGTGTGATC	1235
MCM7	NM 005916	33469967	4176	D-003278-08	GGAGAGAACAAGGATTG	1236
MDM2						
MDM2	NM 002392	4505136	4193	D-003279-05	GGAGATATGTTGTGAAAGA	1237
MDM2	NM 002392	4505136	4193	D-003279-06	CCACAAATCTGATAGTATT	1238
MDM2	NM 002392	4505136	4193	D-003279-07	GATGAGGTATATCAAGTTA	1239
MDM2	NM 002392	4505136	4193	D-003279-08	GGAAGAAACCCAAGACAAA	1240
MKI67						
MKI67	NM 002417	19923216	4288	D-003280-05	GCACAAAGCTTGGTTATAA	1241
MKI67	NM 002417	19923216	4288	D-003280-06	CCTAAGACCTGAACTATTT	1242
MKI67	NM 002417	19923216	4288	D-003280-07	CAAAGAGGAACACAAATTA	1243
MKI67	NM 002417	19923216	4288	D-003280-08	GTAATGGGTCTGTTATTG	1244
MNAT1						
MNAT1	NM 002431	4505224	4331	D-003281-05	GGAAGAAGCTTTAGAAGTG	1245
MNAT1	NM 002431	4505224	4331	D-003281-06	TAGATGAGCTGGAGAGTTC	1246
MNAT1	NM 002431	4505224	4331	D-003281-07	GGACCTTGCTGGAGGCTAT	1247
MNAT1	NM 002431	4505224	4331	D-003281-08	GCAGATAGAGACATATGGA	1248
MYC						
MYC	NM 002467	31543215	4609	D-003282-05	CAGAGAAGCTGGCCTCCTA	1249
MYC	NM 002467	31543215	4609	D-003282-06	GAAACGACGAGAACAGTTG	1250
MYC	NM 002467	31543215	4609	D-003282-07	CGACGAGACCTTCATCAA	1251
MYC	NM 002467	31543215	4609	D-003282-08	CCACACATCAGCACAATA	1252
ORC1L						
ORC1L	NM 004153	31795543	4998	D-003283-05	GAACAGGAATCCAAGACA	1253
ORC1L	NM 004153	31795543	4998	D-003283-06	TAAGAAACGTGCTCGAGTA	1254
ORC1L	NM 004153	31795543	4998	D-003283-07	GAGATCACCTCACCTTCTA	1255
ORC1L	NM 004153	31795543	4998	D-003283-08	GCAGAGAGCCCTTCTTGGA	1256
ORC2L						
ORC2L	NM 006190	32454751	4999	D-003284-05	GAAGAAACCTCCTATGAGA	1257

ORC2L	NM 006190	32454751	4999	D-003284-06	GAAGGGAACTGATGGAGTA	1258
ORC2L	NM 006190	32454751	4999	D-003284-07	GAAGAATGATCCTGAGATT	1259
ORC2L	NM 006190	32454751	4999	D-003284-08	GAAGAGATGTTCAAGAATC	1260
ORC3L						
ORC3L	NM 012381	32483366	23595	D-003285-05	GGACTGCTGTGTAGATATA	1261
ORC3L	NM 012381	32483366	23595	D-003285-06	GAAGTATGACCATACTTG	1262
ORC3L	NM 012381	32483366	23595	D-003285-07	AAAGATCTCTCTGCCAATA	1263
ORC3L	NM 012381	32483366	23595	D-003285-08	CAGCACAGCTAAGAGAATA	1264
ORC4L						
ORC4L	NM 002552	32454749	5000	D-003286-06	GAAAGCACATTCCGTTTAT	1265
ORC4L	NM 002552	32454749	5000	D-003286-07	TGAAAGAAGTCTGGAAT	1266
ORC4L	NM 002552	32454749	5000	D-003286-08	GCTGAGAAGTGGAAATGAA	1267
ORC4L	NM 002552	32454749	5000	D-003286-09	CCAGTGATCTTCATATTAG	1268
ORC5L						
ORC5L	NM 002553	32454752	5001	D-003287-05	GAAATAACCTGTGAAACAT	1269
ORC5L	NM 002553	32454752	5001	D-003287-06	CAGATTACCTCTCTAGTGA	1270
ORC5L	NM 002553	32454752	5001	D-003287-07	GAAGTTCATATTACTCTA	1271
ORC5L	NM 002553	32454752	5001	D-003287-08	GTATTCAGCTGATTTCTAT	1272
ORC6L						
ORC6L	NM 014321	32454755	23594	D-003288-05	GAACATGGCTTCAAAGATA	1273
ORC6L	NM 014321	32454755	23594	D-003288-06	GGACAGGGCTTATTTAATT	1274
ORC6L	NM 014321	32454755	23594	D-003288-07	GAAAGAAGATAGTGGTTGA	1275
ORC6L	NM 014321	32454755	23594	D-003288-08	TATCAGAGCTGTCTTAAAT	1276
PCNA						
PCNA	NM 002592	33239449	5111	D-003289-05	GATCGAGGATGAAGAAGGA	1277
PCNA	NM 002592	33239449	5111	D-003289-07	GCCGAGATCTCAGCCATAT	1278
PCNA	NM 002592	33239449	5111	D-003289-09	GAGGCCTGCTGGGATATTA	1279
PCNA	NM 002592	33239449	5111	D-003289-10	GTGGAGAAGTGGAAATGG	1280
PLK						
PLK	NM 005030	21359872	5347	D-003290-05	CAACCAAAGTCGAATATGA	1281
PLK	NM 005030	21359872	5347	D-003290-06	CAAGAAGAATGAATACAGT	1282
PLK	NM 005030	21359872	5347	D-003290-07	GAAGATGTCCATGGAAATA	1283
PLK	NM 005030	21359872	5347	D-003290-08	CAACACGCCTCATCTCTA	1284
PIN1						
PIN1	NM 006221	5453897	5300	D-003291-05	GGACCAAGGAGGAGGCCCT	1285
PIN1	NM 006221	5453897	5300	D-003291-06	CGTCCTGGCGGCAGGAGAA	1286
PIN1	NM 006221	5453897	5300	D-003291-07	CGGGAGAGGAGGACTTTGA	1287
PIN1	NM 006221	5453897	5300	D-003291-08	AGTCGGGAGAGGAGGACTT	1288
PIN1L						
PIN1L	NM 006222	5453899	5301	D-003292-06	CGACCTGGCGGCAGGAAAT	1289
PIN1L	NM 006222	5453899	5301	D-003292-07	AGGCAGGAGAGAAGGACTT	1290
PIN1L	NM 006222	5453899	5301	D-003292-08	GCTACATCCAGAAGATCAA	1291
PIN1L	NM 006222	5453899	5301	D-003292-09	GGACAGTGTTACGGATTTC	1292
RAD1						
RAD1	NM 002853	19718797	5810	D-003293-05	GAAGATGGACAAATATGTT	1293
RAD1	NM 002853	19718797	5810	D-003293-06	GGAAGAGTCTGTTACTTTT	1294
RAD1	NM 002853	19718797	5810	D-003293-07	GATAACAGAGGCTTCCTTT	1295
RAD1	NM 002853	19718797	5810	D-003293-08	GCATTAGTCTATCTTGTA	1296
RAD17						
RAD17	NM 133338	19718783	5884	D-003294-05	GAATCAAGCTTCCATATGT	1297
RAD17	NM 133338	19718783	5884	D-003294-06	CAACAAAGCCCGAGGATAT	1298
RAD17	NM 133338	19718783	5884	D-003294-07	ACACATGCCTGGAGACTTA	1299
RAD17	NM 133338	19718783	5884	D-003294-08	CTACATAGATTTCTTCATG	1300
RAD9A						
RAD9A	NM 004584	19924112	5883	D-003295-05	TCAGCAAAGTGAATCTTA	1301
RAD9A	NM 004584	19924112	5883	D-003295-06	GACATTGACTCTTACATGA	1302

RAD9A	NM_004584	19924112	5883	D-003295-08	GGAAACCACTATAGGCAAT	1303
RAD9A	NM_004584	19924112	5883	D-003295-09	CGGACGACTTTGCCAATGA	1304
RB1						
RB1	NM_000321	19924112	5925	D-003296-05	GAAAGGACATGTGAACTTA	1305
RB1	NM_000321	19924112	5925	D-003296-06	GAAGAAGTATGATGTATTG	1306
RB1	NM_000321	4506434	5925	D-003296-07	GAAATGACTTCTACTCGAA	1307
RB1	NM_000321	4506434	5925	D-003296-08	GGAGGGAACATCTATATTT	1308
RBBP2						
RBBP2	NM_005056	4826967	5927	D-003297-05	CAAAGAAGCTGAATAAACT	1309
RBBP2	NM_005056	4826967	5927	D-003297-06	CAACACATATGGCGGATTT	1310
RBBP2	NM_005056	4826967	5927	D-003297-07	GGACAAACCTAGAAAGAAG	1311
RBBP2	NM_005056	4826967	5927	D-003297-08	GAAAGGCACTCTCTCTGTT	1312
RBL1						
RBL1	NM_002895	34577078	5933	D-003298-05	CAAGAGAAGTTGTGGCATA	1313
RBL1	NM_002895	34577078	5933	D-003298-06	CAGCAGCACTCCATTTATA	1314
RBL1	NM_002895	34577078	5933	D-003298-07	ACAGAAAGGTCTATCATTT	1315
RBL1	NM_002895	34577078	5933	D-003298-08	GGACATAAAGTTACAATTC	1316
RBL2						
RBL2	NM_005611	21361291	5934	D-003299-05	GAGCAGAGCTTAATCGAAT	1317
RBL2	NM_005611	21361291	5934	D-003299-06	GAGAATAGCCCTTGTGTGA	1318
RBL2	NM_005611	21361291	5934	D-003299-07	GGACTTAGTTTATGGAAAT	1319
RBL2	NM_005611	21361291	5934	D-003299-08	GAATTTAGATGAGCGGATA	1320
RBP1						
RBP1	NM_002899	8400726	5947	D-003300-05	GAGACAAGCTCCAGTGTGT	1321
RBP1	NM_002899	8400726	5947	D-003300-06	GCAAGCAAGTATTCAAGAA	1322
RBP1	NM_002899	8400726	5947	D-003300-07	GCAGGACGGTGACCATATG	1323
RBP1	NM_002899	8400726	5947	D-003300-08	GCAAGTGCATGACAACAGT	1324
RPA3						
RPA3	NM_002947	19923751	6119	D-003322-05	GGAAGTGGTTGGAAGAGTA	1325
RPA3	NM_002947	19923751	6119	D-003322-06	GAAGATAGCCATCCTTTTG	1326
RPA3	NM_002947	19923751	6119	D-003322-07	CATGCTAGCTCAATTCATC	1327
RPA3	NM_002947	19923751	6119	D-003322-08	GATCTTGGACTTTACAATG	1328
SKP1A						
SKP1A	NM_006930	25777710	6500	D-003323-05	GGAGAGATATTTGAAGTTG	1329
SKP1A	NM_006930	25777710	6500	D-003323-06	GGGAATGGATGATGAAGGA	1330
SKP1A	NM_006930	25777710	6500	D-003323-07	CAAACAATCTGTGACTATT	1331
SKP1A	NM_006930	25777710	6500	D-003323-08	TCAATTAAGTTGCAGAGTT	1332
SKP2						
SKP2	NM_005983	16306594	6502	D-003324-05	CATCTAGACTTAAGTGATA	1333
SKP2	NM_005983	16306594	6502	D-003324-06	GAAATCAGATCTCTCTACT	1334
SKP2	NM_005983	16306594	6502	D-003324-07	CTAAAGGTCTCTGGTGTIT	1335
SKP2	NM_005983	16306594	6502	D-003324-08	GATGGTACCCTTCAACTGT	1336
SNK						
SNK	NM_006622	5730054	10769	D-003325-05	GAAGACATCTACAAGCTTA	1337
SNK	NM_006622	5730054	10769	D-003325-06	GAAATACCTTCATGAACAA	1338
SNK	NM_006622	5730054	10769	D-003325-07	GAAGGTCAATGGCTCATAT	1339
SNK	NM_006622	5730054	10769	D-003325-08	CCGGAGATCTCGCGGATTA	1340
STK12						
STK12	NM_004217	4759177	9212	D-003326-07	CAGAAGAGCTGCACATTTG	1341
STK12	NM_004217	4759177	9212	D-003326-08	CCAAACTGCTCAGGCATAA	1342
STK12	NM_004217	4759177	9212	D-003326-09	ACGCGGCACTTCACAATTG	1343
STK12	NM_004217	4759177	9212	D-003326-10	TGGGACACCCGACATCTTA	1344
TFDP1						
TFDP1	NM_007111	34147667	7027	D-003327-05	GGAAGCAGCTCTTGCCAAA	1345
TFDP1	NM_007111	34147667	7027	D-003327-06	GAGGAGACTTGAAAGAATA	1346
TFDP1	NM_007111	34147667	7027	D-003327-07	GAACTTAGAGGTGGAAGA	1347

TFDP1	NM_007111	34147667	7027	D-003327-08	GCGAGAAGGTGCAGAGGAA	1348
TFDP2						
TFDP2	NM_006286	5454111	7029	D-003328-05	GAAAGTGTGTGAGAAAGTT	1349
TFDP2	NM_006286	5454111	7029	D-003328-06	CACAGGACCTTCTTGGTTA	1350
TFDP2	NM_006286	5454111	7029	D-003328-07	CGAAATCCCTGGTGCCAAA	1351
TFDP2	NM_006286	5454111	7029	D-003328-08	TGAGATCCATGATGACATA	1352
TP53						
TP53	NM_000546	8400737	7157	D-003329-05	GAGGTTGGCTCTGACTGTA	1353
TP53	NM_000546	8400737	7157	D-003329-06	CAGTCTACCTCCCGCCATA	1354
TP53	NM_000546	8400737	7157	D-003329-07	GCACAGAGGAAGAGAATCT	1355
TP53	NM_000546	8400737	7157	D-003329-08	GAAGAAACCACTGGATGGA	1356
TP63						
TP63	NM_003722	31543817	8626	D-003330-05	CATCATGTCTGACTATTT	1357
TP63	NM_003722	31543817	8626	D-003330-06	CAAACAAGATTGAGATTAG	1358
TP63	NM_003722	31543817	8626	D-003330-07	GCACACAGACAAATGAATT	1359
TP63	NM_003722	31543817	8626	D-003330-08	CGACAGTCTTGTACAATTT	1360
TP73						
TP73	NM_005427	4885644	7161	D-003331-05	GCAAGCAGCCCATCAAGGA	1361
TP73	NM_005427	4885644	7161	D-003331-06	GAGACGAGGACACGTACTA	1362
TP73	NM_005427	4885644	7161	D-003331-07	CTGCAGAACCTGACCATTG	1363
TP73	NM_005427	4885644	7161	D-003331-08	GGCCATGCCTGTTTACAAG	1364
YWHAZ						
YWHAZ	NM_003406	21735623	7534	D-003332-05	GCAAGGAGCTGAATTATCC	1365
YWHAZ	NM_003406	21735623	7534	D-003332-06	TAAGAGATATCTGCAATGA	1366
YWHAZ	NM_003406	21735623	7534	D-003332-07	GACGGAAGGTGCTGAGAAA	1367
YWHAZ	NM_003406	21735623	7534	D-003332-08	AGAGCAAAGTCTTCTATTT	1368

Table IX

Gene Name	Accession #	GI#	Duplex #	Sequence	SEQ. ID NO.
AR	NM_000044	21322251	D-003400-01	GGAACCTCGATCGTATCATT	1369
AR	NM_000044	21322251	D-003400-02	CAAGGGAGGTTACACCAA	1370
AR	NM_000044	21322251	D-003400-03	TCAAGGAACCTCGATCGTAT	1371
AR	NM_000044	21322251	D-003400-04	GAAATGATTGCACTATTGA	1372
ESR1	NM_000125	4503602	D-003401-01	GAATGTGCCTGGCTAGAGA	1373
ESR1	NM_000125	4503602	D-003401-02	CATGAGAGCTGCCAACCTT	1374
ESR1	NM_000125	4503602	D-003401-03	AGAGAAAGATTGGCCAGTA	1375
ESR1	NM_000125	4503602	D-003401-04	CAAGGAGACTCGCTACTGT	1376
ESR2	NM_001437	10835012	D-003402-01	GAACATCTGCTCAACATGA	1377
ESR2	NM_001437	10835012	D-003402-02	GCACGGCTCCATATACATA	1378
ESR2	NM_001437	10835012	D-003402-03	CAAGAAGATTCCCGGCTTT	1379
ESR2	NM_001437	10835012	D-003402-04	GGAAATGCGTAGAAGGAAT	1380
ESRRA	NM_004451	18860919	D-003403-01	GGCCTTCGCTGAGGACTTA	1381
ESRRA	NM_004451	18860919	D-003403-02	TGAATGCACTGGTGTCTCA	1382
ESRRA	NM_004451	18860919	D-003403-03	GCATTGAGCCTCTCTACAT	1383
ESRRA	NM_004451	18860919	D-003403-04	CCAGACAGCGGGCAAAGTG	1384
ESRRB	NM_004452	22035686	D-003404-01	TACCTGAGCTTACAAATTT	1385
ESRRB	NM_004452	22035686	D-003404-02	GCACCTTCTATAGCGTCAA	1386
ESRRB	NM_004452	22035686	D-003404-03	CAACTCCGATTCCATGTAC	1387
ESRRB	NM_004452	22035686	D-003404-04	GGACTCGCCACCCATGTTT	1388
ESRRG	NM_001438	4503604	D-003405-01	AAACAAAGATCGACACATT	1389

ESRRG	NM 001438	4503604	D-003405-02	TCAGGAACTGTATGATGA	1390
ESRRG	NM 001438	4503604	D-003405-03	GAAGACCAGTCCAAATTAG	1391
ESRRG	NM 001438	4503604	D-003405-04	ATGAAGCGCTGCAGGATTA	1392
HNF4A	NM 000457	21361184	D-003406-01	CGACATCACTGGAGCATAT	1393
HNF4A	NM 000457	21361184	D-003406-02	GAAGGAAGCCGTCCAGAAT	1394
HNF4A	NM 000457	21361184	D-003406-03	CCAAGTACATCCCAGCTTT	1395
HNF4A	NM 000457	21361184	D-003406-04	GGACATGGCCGACTACAGT	1396
HNF4G	NM 004133	6631087	D-003407-01	GCACTGACATAAACGTTAA	1397
HNF4G	NM 004133	6631087	D-003407-02	ACAAAGAGATCCATGATGT	1398
HNF4G	NM 004133	6631087	D-003407-03	AGAGATCCATGATGTATAA	1399
HNF4G	NM 004133	6631087	D-003407-04	AAATGAACGTGACAGAATA	1400
HSAJ2425	NM 017532	8923776	D-003408-01	GAATGAATCTACACCTTTG	1401
HSAJ2425	NM 017532	8923776	D-003408-02	GGAAATACGTGGAGACACT	1402
HSAJ2425	NM 017532	8923776	D-003408-03	CCAGATAACTACGGCGATA	1403
HSAJ2425	NM 017532	8923776	D-003408-04	TGGCGTACCTTCTCATTGA	1404
NR0B1	NM 000475	5016089	D-003409-01	CAGCATGGATGATATGATG	1405
NR0B1	NM 000475	5016089	D-003409-02	CTGCTGAGATTCATCAATG	1406
NR0B1	NM 000475	5016089	D-003409-03	ACAGATTCATCGAACTTAA	1407
NR0B1	NM 000475	5016089	D-003409-04	GAACGTGGCGCTCCTGTAC	1408
NR0B2	NM 021969	13259502	D-003410-01	GAATATGCCTGCCTGAAAG	1409
NR0B2	NM 021969	13259502	D-003410-02	GGAATATGCCTGCCTGAAA	1410
NR0B2	NM 021969	13259502	D-003410-03	CGTAGCCGCTGCCTATGTA	1411
NR0B2	NM 021969	13259502	D-003410-04	GCCATTCTCTACGCACTTC	1412
NR1D1	NM 021724	13430847	D-003411-01	CAACACAGGTGGCGTCATCTT	1413
NR1D1	NM 021724	13430847	D-003411-02	GGCATGGTGTTACTGTGTATT	1414
NR1D1	NM 021724	13430847	D-003411-03	CAACATGCATTCCGAGAAGTT	1415
NR1D1	NM 021724	13430847	D-003411-04	GCGCTTGCTTCGTTGTTCTT	1416
NR1H2	NM 007121	11321629	D-003412-01	GAACAGATCCGGAAGAAGA	1417
NR1H2	NM 007121	11321629	D-003412-02	GAAGAACAGATCCGGAAGA	1418
NR1H2	NM 007121	11321629	D-003412-03	CTAAGCAAGTGCCTGGTTT	1419
NR1H2	NM 007121	11321629	D-003412-04	GCTAACAGCGGCTCAAGAA	1420
NR1H3	NM 005693	5031892	D-003413-01	GAACAGATCCGCCTGAAGA	1421
NR1H3	NM 005693	5031892	D-003413-02	GGAGATAGTTGACTTTGCT	1422
NR1H3	NM 005693	5031892	D-003413-03	GAGTTTGCCTTGCTCATTG	1423
NR1H3	NM 005693	5031892	D-003413-04	TGACTTTGCTAAACAGCTA	1424
NR1H4	NM 005123	4826979	D-003414-01	CAAGTGACCTCGACAACAA	1425
NR1H4	NM 005123	4826979	D-003414-02	GAAAGAATTGAAATAGTG	1426
NR1H4	NM 005123	4826979	D-003414-03	CAACAGACTCTTCTACATT	1427
NR1H4	NM 005123	4826979	D-003414-04	GAACCATACTCGCAATACA	1428
NR1I2	NM 003889	11863133	D-003415-01	GAACCATGCTGACTTTGTA	1429
NR1I2	NM 003889	11863133	D-003415-02	GATGGACGCTCAGATGAAA	1430
NR1I2	NM 003889	11863133	D-003415-03	CAACCTACATGTTCAAAGG	1431
NR1I2	NM 003889	11863133	D-003415-04	CAGGAGCAATTGCGCATTA	1432
NR1I3	NM 005122	4826660	D-003416-01	GGAAATCTGTCACATCGTA	1433
NR1I3	NM 005122	4826660	D-003416-02	TCGCAGACATCAACACTTT	1434

NR1I3	NM 005122	4826660	D-003416-03	CCTCTTCGCTACACAATTG	1435
NR1I3	NM 005122	4826660	D-003416-04	GAACAGTTTGTGCAGTTTA	1436
NR2C1	NM 003297	4507672	D-003417-01	TGACAGCACTTGATCATAA	1437
NR2C1	NM 003297	4507672	D-003417-02	GGAAGGAAGTGTACACCTA	1438
NR2C1	NM 003297	4507672	D-003417-03	GAGCACATCTTCAAACCTAC	1439
NR2C1	NM 003297	4507672	D-003417-04	GAAGAAATTGCACATCAAA	1440
NR2C2	NM 003298	4507674	D-003418-01	GAACAACGGTGACACTTCA	1441
NR2C2	NM 003298	4507674	D-003418-02	CTGATGAGCTCCAACATAA	1442
NR2C2	NM 003298	4507674	D-003418-03	CAACCTAAGTGAATCTTTG	1443
NR2C2	NM 003298	4507674	D-003418-04	GAAGACACCTACCGATTGG	1444
NR2E1	NM 003269	21361108	D-003419-01	GATCATATCTGAAATACAG	1445
NR2E1	NM 003269	21361108	D-003419-02	CAAGACTGCTTTCAGATAT	1446
NR2E1	NM 003269	21361108	D-003419-03	GTTAGATGCTACTGAATTT	1447
NR2E1	NM 003269	21361108	D-003419-04	CAATGTATCTCTATGAAGT	1448
NR2E3	NM 014249	7657394	D-003420-01	GAGAAGCTCCTTTGTGATA	1449
NR2E3	NM 014249	7657394	D-003420-02	GAAGCACTATGGCATCTAT	1450
NR2E3	NM 014249	7657394	D-003420-03	GAAGGATCCTGAGCACGTA	1451
NR2E3	NM 014249	7657394	D-003420-04	GAAGCTCCTTTGTGATATG	1452
NR2F1	NM 005654	20127484	D-003421-01	GAAACTCTCATCCGCGATA	1453
NR2F1	NM 005654	20127484	D-003421-02	TCTCATCCGCGATATGTTA	1454
NR2F1	NM 005654	20127484	D-003421-03	CAAGAAAGTGCCTCAAAGTG	1455
NR2F1	NM 005654	20127484	D-003421-04	GGAACCTTAACCTACACATG	1456
NR2F2	NM 021005	14149745	D-003422-01	GTACCTGTCCGGATATATT	1457
NR2F2	NM 021005	14149745	D-003422-02	CCAACCAGCCGACGAGATT	1458
NR2F2	NM 021005	14149745	D-003422-03	ACTCGTACCTGTCCGATA	1459
NR2F2	NM 021005	14149745	D-003422-04	GGCCGTATATGGCAATTCA	1460
NR2F6	NM 005234	20070198	D-003423-01	CGACGCCTGTGGCCTCTCA	1461
NR2F6	NM 005234	20070198	D-003423-02	CAGCCGGTGTCCGAACCTGA	1462
NR2F6	NM 005234	20070198	D-003423-03	CAACCGTGAAGTCCAGATC	1463
NR2F6	NM 005234	20070198	D-003423-04	GTACTGCCTGTCTCAAGAAG	1464
NR3C1	NM 000176	4504132	D-003424-01	GAGGACAGATGTACCACTA	1465
NR3C1	NM 000176	4504132	D-003424-02	GATAAGACCATGAGTATTG	1466
NR3C1	NM 000176	4504132	D-003424-03	GAAGACGATTCATTCTTT	1467
NR3C1	NM 000176	4504132	D-003424-04	GGACAGATGTACCACTATG	1468
NR3C2	NM 000901	4505198	D-003425-01	GCAAACAGATGATCCAAGT	1469
NR3C2	NM 000901	4505198	D-003425-02	CAGCTAAGATTTATCAGAA	1470
NR3C2	NM 000901	4505198	D-003425-03	GCACGAAAGTCAAAGAAGT	1471
NR3C2	NM 000901	4505198	D-003425-04	GGTATCCGGTCTTAGAATA	1472
NR4A1	NM 002135	21361341	D-003426-01	GAAGGAAGTTGTCCGAACA	1473
NR4A1	NM 002135	21361341	D-003426-02	CAGGAGAGTTTGACACCTT	1474
NR4A1	NM 002135	21361341	D-003426-03	CAGTGGCTCTGACTACTAT	1475
NR4A1	NM 002135	21361341	D-003426-04	GAAGGCCGCTGTGCTGTGT	1476
NR4A2	NM 006186	5453821	D-003427-01	GCAATGCGTTCGTGGCTTT	1477
NR4A2	NM 006186	5453821	D-003427-02	CGGCTACACAGGAGAGTTT	1478
NR4A2	NM 006186	5453821	D-003427-03	CCACGTGACTTTCAACAAT	1479

NR4A2	NM 006186	5453821	D-003427-04	GAATACAGCTCCGATTTCT	1480
NR4A3	NM 006981	11276070	D-003428-01	CAAAGAAGATCAGACATTA	1481
NR4A3	NM 006981	11276070	D-003428-02	GATCAGACATTACTTATTG	1482
NR4A3	NM 006981	11276070	D-003428-03	CCAGAGATCTTGATTATTC	1483
NR4A3	NM 006981	11276070	D-003428-04	GAAGTTGTCCGTACAGATA	1484
NR5A1	NM 004959	20070192	D-003429-01	GATTTGAAGTTCCTGAATA	1485
NR5A1	NM 004959	20070192	D-003429-02	GGAGCGAGCTGCTGGTGTT	1486
NR5A1	NM 004959	20070192	D-003429-03	GGAGGTGGCCGACCAGATG	1487
NR5A1	NM 004959	20070192	D-003429-04	CAACGTGCCTGAGCTCATC	1488
NR5A2	NM 003822	20070161	D-003430-01	CCAAACATATGGCCACTTT	1489
NR5A2	NM 003822	20070161	D-003430-02	TCAGAGAACTTAAGGTTGA	1490
NR5A2	NM 003822	20070161	D-003430-03	GGATCCATCTTCCTGGTTA	1491
NR5A2	NM 003822	20070161	D-003430-04	AAGAATACCTCTACTACAA	1492
NR6A1	NM 033334	15451847	D-003431-01	CAACGAACCTGTCTCATTT	1493
NR6A1	NM 033334	15451847	D-003431-02	GAAGAACTACACAGATTTA	1494
NR6A1	NM 033334	15451847	D-003431-03	GAAGATGGATACGCTGTGA	1495
NR6A1	NM 033334	15451847	D-003431-04	AAACGATACTGGTACATTT	1496
null	D16815	2116671	D-003432-01	GAAGAATGATCGAATAGAT	1497
null	D16815	2116671	D-003432-02	GAACATGGAGCAATATAAT	1498
null	D16815	2116671	D-003432-03	GAGGAGCTCTTGCCCTTTA	1499
null	D16815	2116671	D-003432-04	TAAACAACATGCACTCTGA	1500
PGR	NM 000926	4505766	D-003433-01	GAGATGAGGTCAAGCTACA	1501
PGR	NM 000926	4505766	D-003433-02	CAGCGTTTCTATCAACTTA	1502
PGR	NM 000926	4505766	D-003433-03	AGATAACTCTCATTCAGTA	1503
PGR	NM 000926	4505766	D-003433-04	GTAGTCAAGTGGTCTAAAT	1504
PPARA	NM 005036	7549810	D-003434-01	TCACGGAGCTGACGGAATT	1505
PPARA	NM 005036	7549810	D-003434-02	GAACATGACATAGAAGATT	1506
PPARA	NM 005036	7549810	D-003434-03	GGATAGTTCTGGAAGCTTT	1507
PPARA	NM 005036	7549810	D-003434-04	GACTCAAGCTGGTGTATGA	1508
PPARD	NM 006238	5453939	D-003435-01	GAGCGCAGCTGCAAGATTC	1509
PPARD	NM 006238	5453939	D-003435-02	GCATGAAGCTGGAGTACGA	1510
PPARD	NM 006238	5453939	D-003435-03	GGAAGCAGTTGGTGAATGG	1511
PPARD	NM 006238	5453939	D-003435-04	GCTGCAAGATTGAGAAGAA	1512
PPARG	NM 138712	20336234	D-003436-01	AGACTCAGCTCTACAATAA	1513
PPARG	NM 138712	20336234	D-003436-02	GATTGAAGCTTATCTATGA	1514
PPARG	NM 138712	20336234	D-003436-03	AAGTAACTCTCCTCAAATA	1515
PPARG	NM 138712	20336234	D-003436-04	GCATTTCTACTCCACATTA	1516
RARA	NM 000964	4506418	D-003437-01	GACAAGAAGTGCATCATCA	1517
RARA	NM 000964	4506418	D-003437-02	GCAAAATACACTACGAACAA	1518
RARA	NM 000964	4506418	D-003437-03	GAACAACAGCTCAGAACAA	1519
RARA	NM 000964	4506418	D-003437-04	GAGCAGCAGTTCTGAAGAG	1520
RARB	NM 000965	14916493	D-003438-01	GCACACTGCTCAATCAATT	1521
RARB	NM 000965	14916493	D-003438-02	GCAGAAGTATTGAGAAGAA	1522
RARB	NM 000965	14916493	D-003438-03	GGAATGACAGGAACAAGAA	1523
RARB	NM 000965	14916493	D-003438-04	GCACAGTCCTAGCATCTCA	1524

RARG	NM 000966	21359851	D-003439-01	GAAATGACCGGAACAAGAA	1525
RARG	NM 000966	21359851	D-003439-02	TAGAAGAGCTCATCACCAA	1526
RARG	NM 000966	21359851	D-003439-03	CAAGGAAGCTGTGCGAAAT	1527
RARG	NM 000966	21359851	D-003439-04	TCAGTGAGCTGGCTACCAA	1528
RORA	NM 134261	19743902	D-003440-01	GGAAAGAGTTTATGTTCTA	1529
RORA	NM 134261	19743902	D-003440-02	CAAGATCTGTGGAGACAAA	1530
RORA	NM 134261	19743902	D-003440-03	GCACCTGACTGAAGATGAA	1531
RORA	NM 134261	19743902	D-003440-04	CCGAGAAGATGGAATACTA	1532
RORB	NM 006914	19743906	D-003441-01	GCACAGAACATCATTAAAGT	1533
RORB	NM 006914	19743906	D-003441-02	CCACACCTATGAAGAAATT	1534
RORB	NM 006914	19743906	D-003441-03	GATCAAATTCTACTTCTGA	1535
RORB	NM 006914	19743906	D-003441-04	TCAAACAGATAAAGCAAGA	1536
RORC	NM 005060	19743908	D-003442-01	TAGAACAGCTGCAGTACAA	1537
RORC	NM 005060	19743908	D-003442-02	TCACCGAGGCCATTAGTA	1538
RORC	NM 005060	19743908	D-003442-03	GAACAGCTGCAGTACAATC	1539
RORC	NM 005060	19743908	D-003442-04	CCTCATGCCACCTTGAATA	1540
RXRA	NM 002957	21536318	D-003443-01	TGACGGAGCTTGTGTCCAA	1541
RXRA	NM 002957	21536318	D-003443-02	CAACAAGGACTGCCTGATT	1542
RXRA	NM 002957	21536318	D-003443-03	GCAAGGACCTGACCTACAC	1543
RXRA	NM 002957	21536318	D-003443-04	GCAAGGACCGGAACGAGAA	1544
RXRB	NM 021976	21687229	D-003444-01	GCAAAGACCTTACATACTC	1545
RXRB	NM 021976	21687229	D-003444-02	GCAATCATTCTGTTTAATC	1546
RXRB	NM 021976	21687229	D-003444-03	TCACACCGATCCATTGATG	1547
RXRB	NM 021976	21687229	D-003444-04	GCAAACGGCTATGTGCAAT	1548
RXRG	NM 006917	21361386	D-003445-01	GGAAGGACCTCATCTACAC	1549
RXRG	NM 006917	21361386	D-003445-02	CCGGATCTCTGGTTAAACA	1550
RXRG	NM 006917	21361386	D-003445-03	GCGAGCCATTGTACTCTTT	1551
RXRG	NM 006917	21361386	D-003445-04	GAGCCATTGTACTCTTTAA	1552
THRA	NM 003250	20127451	D-003446-01	GGACAAAGACGAGCAGTGT	1553
THRA	NM 003250	20127451	D-003446-02	GGAAACAGAGGCGGAAATT	1554
THRA	NM 003250	20127451	D-003446-03	GTAAGCTGATTGAGCAGAA	1555
THRA	NM 003250	20127451	D-003446-04	GAACCTCCATCCCACCTAT	1556
THRB	NM 000461	10835122	D-003447-01	GAATGTCGCTTTAAGAAAT	1557
THRB	NM 000461	10835122	D-003447-02	GAACAGTCGTCGCCACATC	1558
THRB	NM 000461	10835122	D-003447-03	GGACAAGCACCAATAGTCA	1559
THRB	NM 000461	10835122	D-003447-04	GTGGAAAGGTTGACTTGGA	1560
VDR	NM 000376	4507882	D-003448-01	TGAAGAAGCTGAACTTGCA	1561
VDR	NM 000376	4507882	D-003448-02	GCAACCAAGACTACAAGTA	1562
VDR	NM 000376	4507882	D-003448-03	TCAATGCTATGACCTGTGA	1563
VDR	NM 000376	4507882	D-003448-04	CCATTGAGGTCATCATGTT	1564

Table X

Gene Symbol	Sense	SEQ ID NO
ABCB1	GACCAUAAAUGUAAGGUUU	1565
	UAGAAGAUCUGAUGUCAA	1566
	GAAUGUUCACUUCAGUUA	1567
	GAAGAUCGCUACUGAAGCA	1568
ABCC1	GGAAGCAACUGCAGAGACA	1569
	GAUGACACCUCUCAACAAA	1570
	UAAAGUUGCUCUCAAGUU	1571
	CAACGAGUCUGCCGAAGGA	1572
ABCG2	GCAGAUGCCUUCUUCGUUA	1573
	AGGCAAAUCUUCGUUAUUA	1574
	GGGAAGAAUCUGGUCUAA	1575
	UGACUCAUCCCAACAUUUA	1576
KCNH2	CCGACGUGCUGCCUGAGUA	1577
	GAGAAGAGCAGCGACACUU	1578
	GAUCAUAGCACCUAAGAU	1579
	GCUAUUUACUGCUCUUAUU	1580
	UCACUGGGCUCUUAUUAUU	1581
	GUGCGAGCCUUCUGAAUAU	1582
	GCUAAGCUAUACUACUGUA	1583
	UGACGGCGCUCUACUUCAC	1584
KCNH1	GAGAUGAAUCCUUGAAA	1585
	GAAGAACGCAUGAAACGAA	1586
	GAUAAAGACACGAUUGAAA	1587
	GCUGAGAGGUCUAUUUAAA	1588
CLCA1	GAACAACAAUGGCUAUGAA	1589
	GUACAUACGUGGUGGAUU	1590
	GAACAGCUCACAAGUAUAU	1591
	GGAAACGUGUGUCUAUAUU	1592
SLC6A1	GGAGGUGGGAGGACAGUUA	1593
	UCACAGCCUGGUGGAUGA	1594
	GAAGCUGGCUCCUAUGUUC	1595
	GGUCAACACUACCAACAUG	1596
SLC6A2	GAACACAAGGUCAACAUUG	1597
	AGAAGGAGCUGGCCUAGUG	1598
	CGGAAACUCUUCACAUUUG	1599
	CAACAAAUUUGACAACAAC	1600
SLC21A2	GUACAUCUCCAUCUUAUUU	1601
	GGAAGUGGCUGAGUUAUUA	1602

GAAGGGAGGCUCAAUGUAA	1603
GAAGGAAGUGGCUGAGUUA	1604

SLC21A3	Sense	
	GUAGAAACAGGAGCUAUUA	1605
	CAAGAUUACUGUCAAAACA	1606
	GCACAAGAGUAUUUGGUAA	1607
	GCAAAUGUCCCUUCUGUAU	1608
	GCAUGACUCCUAUAUAAUA	1609
	AAACAGCAAUUUCCCUUAA	1610
	GAAAUGCCUCUUCAGGAA	1611

SLC28A1	Sense	
	GUUCAUCGCUCUCCUCUUU	1612
	GGAUCAAGCUGUUUCUGAA	1613
	GGACUGCAGUUUGUACUUG	1614
	GAGUGAAACUGACCUAUGG	1615

SLC29A1	Sense	
	GAACGCUGCUCGUGGAA	1616
	GAAAGCCACUCUAUCAAAG	1617
	GAAACCAGGUGCCUUCAGA	1618
	CCUCACAGCUGUAUUAUG	1619

SLC26A1	Sense	
	CCACGGAGCUGCUGGUCAU	1620
	GGGUUGACAUCUUAUUUGA	1621
	GCACGAGGGUCUCUGUGUU	1622
	GGCCAUCGCCUACUCAUUG	1623
	CAACACCCAUGGCAAUUA	1624
	GAGGAAAGAUUCUUGCUGAU	1625
	GAGCAAGCGUCCUCCAAAU	1626
	GCAACACCCAUGGCAAUUA	1627

SLC26A2	Sense	
	CCAAAGAACUCAAUAGAACA	1628
	ACAAGAACCUUCAGACUAA	1629
	GAAGGUAGAUAGAAGAAUG	1630
	GUAUUGAACUGUACUGUAA	1631

SLC4A4	Sense	
	GCAAUUCUCUUAUUUAUC	1632
	GGAAAGAUGUCCACUGAAA	1633
	GGACAAAGCCUUCUCAAU	1634
	GGAAUGGGAUCCAGCAAUU	1635

GLRA1	Sense	
	UGAAAGCCAUUGACAUUUG	1636
	CAGACACGCUGGAGUUUAA	1637
	CAAUAGCGCUUUCUGGUUU	1638
	GCAGGUAGCAGAUUGGACUA	1639

KLK1	Sense	
	UCAGAGUGCUGUCUUAUGU	1640
	CAACUUGUUUGACGACGAA	1641
	UGACAGAGCCUGCUGAUAC	1642

	AGGCGGCUCUGUACCAUUU	1643
ADAM2	Sense	
	GAAACAUGCUGUGAUUUUG	1644
	GCAGAUGUUUCCUUAUAUA	1645
	CAACAGAGAUGCCAUGAUA	1646
	GAAAGGCGCUACAUUGAGA	1647
XPNPE P1	Sense	
	GACCUGAGCUUCCCAACAA	1648
	GCGACUGGCUCAACAAUUA	1649
	GAGAUUGCGUGGCUAUUUA	1650
	GACAGCAACUGGACACUUA	1651
GZMA	Sense	
	GGAAGAGACUCGUGCAAUG	1652
	GGAACCAUGUGCCAAGUUG	1653
	GAAGUAACUCCUCAUUCAA	1654
	GAACUCCUAUAGAUUUCUG	1655
CMKLR1	Sense	
	CAUAGAAGCUUUACCAAGA	1656
	GAAUGGAGGAUGAAGAUUA	1657
	GGUCAAUUGCUCUAAGUGAA	1658
	GAGAGGACUUCUAUGAAUG	1659
CLN3	Sense	
	CAUCAUGCCUUCUGAAUAA	1660
	CAACAGCUCUAUCACGAUUU	1661
	GCAACAACUUCUCUUAUGU	1662
	GGUCUUCGCUAGCAUCUCA	1663
CALCR	Sense	
	GGACCUAGCUGUUGUAAAAG	1664
	GAAAGACCAUGCAUUUAAA	1665
	GCAGGAAGAUGUAUGCUUU	1666
	GAAUAAACCAGUAUUGUUA	1667
OXTR	Sense	
	GGACCCAGAUAUCCAAUA	1668
	GCAAUACUAUCCUAACUGA	1669
	GAAUAUAGAUUAGCGUUUG	1670
	GAUGAGGCAUGACUACUAA	1671
EDG4	Sense	
	GCGAGUCUGUCCACUAUAC	1672
	GAGAACGGCCACCCACUGA	1673
	GAACGGCCACCCACUGAUG	1674
	GGUCAAUUGCUGCUGUGUAC	1675
EDG5	Sense	
	UCCAGGAACACUAUAAUUA	1676
	GUGACCAUCUUCUCCAUCA	1677
	CAUCCUCUGUUGCGCCAUAU	1678
	CCAACAAGGUCCAGGAACA	1679

EDG7	Sense	
	ACACUGAUACUGUCGAUGA	1680
	AAUAGGAGCAACACUGAUA	1681
	CAGCAGGAGUUACCUUGUU	1682
	GGACACCCAUGAAGCUAAU	1683
PTCH	Sense	
	GCACAGAACUCCACUCAA	1684
	GGACAGCAGUUCAUUGUUA	1685
	GAGAAGAGGCCUAUGUUUAA	1686
	GGACAAACUUCGACCCUUU	1687
SMO	Sense	
	UCGCUACCCUGCUGUUAUU	1688
	GCUACAAGAACUACCGAUA	1689
	CAAGAAAGCUUCCUUAAC	1690
	GAGAAGAAUACAGUCAAU	1691
CASP3	Sense	
	CAAUUAUCUGAAGAGCUA	1692
	GAACUGGACUGUGGCAUUG	1693
	GUGAGAAGAUGGUUAUUUU	1694
	GAGGGUACUUUAAGACAU	1695
CASP6	Sense	
	CAUGAGGUGUCAACUGUUA	1696
	GAAGUGAAAUGCUUUAUG	1697
	AAAUUUGGCCUCCUCCUAG	1698
	GCAAUCACAUUUAUGCAUA	1699
	CAACAUAAACUGAGGUGGAU	1700
	CAUGGUACAUUCAAGAUUU	1701
CASP7	Sense	
	GAACUCUACUUCAGUCAAU	1704
	GGGCAAAUGCAUCAUAAUA	1703
	CAACAGAGGGAGUUUAAUA	1704
	GAACAAAGCCACUGACUGA	1705
CASP8	Sense	
	GAAGUGAACUAUGAAGUAA	1706
	CAACAAGGAUGACAAGAAA	1707
	GGACAAAGUUUACCAAUG	1708
	GAGGGUCGAUCAUCUAUUA	1709
	GAAUAUAGAGGGCUUAUGA	1710
	CAACGACUAUGAAGAAUUC	1711
	GAAGUGAGCAGAUCAAGAU	1712
	GAGGAAAUCUCCAAUUGCA	1713
CASP9	Sense	
	CCAGGCAGCUGAUCAUAGA	1714
	UCUCAGGUGUUGCCAAUA	1715
	GAACAGCUGUAAUCUAUGA	1716
	CCACUGGUCUGUAGGGAUU	1717
DVL1	Sense	

UCGUAAAGCUGUUGAUUAUC	1718
GAGGAGAUCUUUGAUGACA	1719
GUAAAGCUGUUGAUUAUCGA	1720
GAUCGUAAAGCUGUUGAUA	1721

DVL2	Sense	
	AGACGAAGGUGAUUUACCA	1722
	UGUGAGAGCUACCUAGUCA	1723
	GAAGAAAUUUCAGAUGACA	1724
	UAAUAGGCAUUUCCUCUUU	1725

PTEN	Sense	
	GUGAAGAUCUUGACCAAUG	1726
	GAUCAGCAUACACAAAUUA	1727
	GAAUGAACCUUCUGCAACA	1728
	GGCGCUAUGUGUAUUUAUA	1729

PDK1	Sense	
	GUACAAAGCUGGUUAUAUCC	1730
	GAAAGACUCCCAGUGUAUA	1731
	GGAAGUCCAUCUCAUCGAA	1732
	CCAAAGACAUGACGACGUU	1733

PDK2	Sense	
	GUAAAGAGGAGACUGAAUG	1734
	GGUCUGUGAUGGUCCCUAA	1735
	CAAAGAUGCCUACGACAUG	1736
	GGGCGAUGCCUGAGGGUUA	1737

PPP2CA	Sense	
	UCACACAAGUUUAUGGUUU	1738
	CAACAGCCGUGACCACUUU	1739
	UAACCAAGCUGCAAUCAUG	1740
	GAACUUGACGAUACUCUAA	1741

CTNNA1	Sense	
	GAAGAGAGGUCGUUCUAAG	1742
	AAGCAGAUGUGCAUGAUUA	1743
	UCUAAUAACUGCAGUGUUU	1744
	GUAAAGGGCCCUCUAAUAA	1745

CTNNA2	Sense	
	GAAAGAAUAUGCCCAAGUU	1746
	GAAGAAGAAUGCCACAAUG	1747
	GCAGGAAGAUUAUGAUGUG	1748
	AAAGAAAGCCCAUGUACUA	1749

HSPCA	Sense	
	GGGAAAGAGCUGCAUAUUA	1750
	GCUUAGAACUCUUUACUGA	1751
	UAUAAGAGCUUGACCAAUG	1752
	GCAGAUUAUCUCUAUGAUUG	1753

DCTN2	Sense	
	CAACUCAUGUCCAAUACUG	1754
	GGAAUGAGCCAGAUGUUUA	1755

	GGAGACAGCUGUACGUUGU	1756
	UCCAAGAGCUGACAACUGA	1757
CD2	Sense	
	GUAAGGAGAAGCAAUAUAA	1758
	AAGAUGAGCUUCCAUGUA	1759
	GGACAUCUAUCUCAUCAUU	1760
	GACAAGAGCCCACAGAGUA	1761
BAD	Sense	
	GUACUUCCCUCAGGCCUAU	1762
	GCUGUGCCUUGACUACGUA	1763
	GUACUUCCCUCAGGCCUAU	1764
	GGUCAGGUGCCUCGAGAUC	1765
SMAC	Sense	
	CAGCGUAACUUCAUUCUUC	1766
	UAACUUCAUUCUUCAGGUA	1767
	CAGCUGCUCUUACCCAUUU	1768
	GAUUGAAGCUAUUACUGAA	1769
	UAGAAGAGCUCCGUCAGAA	1770
	CCACAU AUGCGUUGAUUGA	1771
	GCGCAGGGCUCUCUACCUA	1772
MAP3K5	Sense	
	GAACAGCCUUCAAAUCAAA	1773
	GAUGUUCUCUACUAUGUUA	1774
	GCAAAUACUGGAAGGAUUA	1775
	CAGGAAAGCUCGUAAUUA	1776
PVR	Sense	
	CCACACGGCUGACCUCAUA	1777
	CAGCAGAAUCCUCUUAUA	1778
	GCAGAAUCCUCUUAUAAA	1779
	GAUCGGGAUUUAUUUCUAU	1780
ERBB2	Sense	
	UGUGGGAGCUGAUGACUUU	1781
	UCACAGAGAUCUUGAAAGG	1782
	UGGAAGAGAUACAGGUUA	1783
	GCUCAUCGCUCACAACCAA	1784
SOS1	Sense	
	GAGCACCACUUCUAUGAUU	1785
	CAAAGAAGCUGUCAAUAU	1786
	UGAAAGCCCUCUUUAUUA	1787
	GAAAUAGCAUGGAGAAGGA	1788
BRCA1	Sense	
	CCAUACAGCUUCAUAAUA	1789
	GAAGAGAACUUAUCUAGUG	1790
	GAAGUGGGCUCCAGUAUUA	1791
	GCAAGAUGCUGAUUCAUUA	1792
	GAAGUGGGCUCCAGUAUUA	1793
	GAACGGACACUGAAUAUU	1794
	GCAGAUAGUUCUACCAGUA	1795

CDKN1A	Sense	
	GAACAAGGAGUCAGACAUU	1796
	AAACUAGGCGGUUGAAUGA	1797
	GAUGGAACUUCGACUUUGU	1798
	GUAAACAGAUGGCACUUUG	1799
CDKN1B	Sense	
	GGAAUGGACAUCCUGUAUA	1800
	GGAGAAAGAUGUCAAACGU	1801
	GAAUGGACAUCCUGUAUAA	1802
	GUAAACAGCUCGAAUUAAG	1803
SLC2A4	Sense	
	CAGAUAGGCUCCGAAGAUG	1804
	AGACUCAGCUCAGAAUAC	1805
	GAUCGGUUCUUUCAUCUUC	1806
	CAGGAUCGGUUCUUUCAUC	1807
NOS2A	Sense	
	CCAGAUAAAGUGACAUAAAGU	1808
	UAAGUGACCUGCUUUGUAA	1809
	GAAGAGAGAUUCCAUUGAA	1810
	UGAAAGAGCUCACAAACAA	1811
FRAP1	Sense	
	GAGCAUGCCGUCAAUAAUA	1812
	CAAGAGAACUCAUAUAAG	1813
	CCAAAGUGCUGCAGUACUA	1814
	UAAGAAAGCUAUCCAGAUU	1815
FKBP1A	Sense	
	GAAACAAGCCCUUUAAGUU	1816
	GAAUUACUCUCCAAGUUGA	1817
	CAGCACAAGUGGUAGGUUA	1818
	GUUGAGGACUGAAUUACUC	1819
	GAUGGCAGCUGUUUAAAUG	1820
	GAGUAUCCUUUCAGUGUUA	1821
TNFRSF 1A	Sense	
	CAAAGGAACCUACUUGUAC	1822
	GGAACCUACUUGUACAAUG	1823
	GAACCUACUUGUACAAUGA	1824
	GAGUGUGUCUCCUGUAGUA	1825
IL1R1	Sense	
	GGACAAGAAUCAUUGGAUA	1826
	GAACAAGCCUCCAGGAUUC	1827
	GGACUUGUGUGCCCUUAUA	1828
	GAACACAAAGGCACUAUAA	1829
IRAK1	Sense	
	CGAAGAAAGUGAUGAAUUU	1830
	GCUCUUUGCCCAUCUCUUU	1831
	UGAAAGACCUGGUGGAAGA	1832
	GCAAUUCAGUUUCUACAUC	1833

TRAF2	Sense	
	GAAGACAGAGUUAUUAAC	1834
	UCACGAAGACAGAGUUAU	1835
	AGACAGAGUUAUUAACCA	1836
	CACGAAGACAGAGUUAUA	1837
	GCUGAAGCCUGUCUGAUGU	1838

TRAF6	Sense	
	CAAUUGAUCUGAGGCAGUU	1839
	GUUCAUAGUUUGAGCGUUA	1840
	GGAGAAACCUGUUGUGAUU	1841
	GGACAAAGUUGCUGAAAUC	1842
	CAAUUGAUCUGAGGCAGUU	1843
	GGAGAAACCUGUUGUGAUU	1844
	GGACAAAGUUGCUGAAAUC	1845
	GUUCAUAGUUUGAGCGUUA	1846

TRADD	Sense	
	UGAAGCACCUUGAUCUUUG	1847
	GGGAGCGCAUACCUGUUU	1848
	GAGGAGCGCUGUUUGAGUU	1849
	GGACGAGGAGCGCUGUUUG	1850
	GAGGAGCGCUGUUUGAGUU	1851
	GAUGUCUCUCUCCUCUUU	1852
	GCUCACUCCUUUCUACUAA	1853
	UGAAGCACCUUGAUCUUUG	1854

FADD	Sense	
	GCACAGAUUUUCCAUUUC	1855
	GCAGUCCUCUUAUUCUAA	1856
	GAACUCAAGCUGCGUUUAU	1857
	GGACGAAUUGAGAUAAUAU	1858

IKBKE	Sense	
	UAAGAACACUGCUCAUGAA	1859
	GAGGCAUCCUGAAGCAUUA	1860
	GAAGGCGGCUGCAGAACUG	1861
	GGAACAAGGAGAUCAUGUA	1862

IKBKG	Sense	
	CUAUCGAGGUCGUUAAAUU	1863
	GAAUGCAGCUGGAAGAUCU	1864
	GCGGCGAGCUGGACUGUUU	1865
	CCAGACCGAUGUGUAUUUA	1866

TNFRSF 5	Sense	
	GGUCUCACCUCGCUAUGGU	1867
	GAAAGCGAAUCCUAGACA	1868
	GCACAAACAAGACUGAUGU	1869
	GAAGGGCACCUCAGAAACA	1870
	UCUCCCAACUUGUAUUAAA	1871

RELA	Sense	
	UCAAGUGUCUCCAUCAUG	1872
	UCAAGUGCCUUAUAGUAG	1873

	GGAGUACCCUGAGGCUAUA	1874
	GAUGAGAUCUCCUACUGU	1875
ARHA	Sense	
	GAGCUGGGCUAAGUAAUA	1876
	GACCAAAGAUGGAGUGAGA	1877
	GGAAGAAACUGGUGAUUGU	1878
	GGCUGUAACUACUUUAUA	1879
CDC42	Sense	
	GGACAUUUGUUUGCCAUUU	1880
	GGAGAACCAUAUACUCUUG	1881
	GAACCAUUGCUIUCUCAUG	1882
	GAAGACCUGUUAUGUAGAG	1883
	GAUCAAGAAUUGCAAUAUC	1884
	GAAAAGGGGUGACCUAGUA	1885
	UGACAAACCUUAUGGAAA	1886
ROCK1	Sense	
	GGAUGAGCUUCAGAUGCA	1887
	GGACACAGCUGUAAGAUUG	1888
	GACAAGAGAUUACAGAUAA	1889
	GAAGAAACAUUCCCUAUUC	1890
PAK1	Sense	
	GAGGGUGGUUUAUGAUUAA	1891
	CAACAAAGAACAUCACUA	1892
	GAAGAAUUAUACACGGUUU	1893
	UACAUGAGCUUACAGUA	1894
PAK2	Sense	
	GGUAGGAGAUGAAUUGUUU	1895
	AGAAGGAACUGAUCAUUA	1896
	CUACAGACCUCCAUAUCA	1897
	GAAACUGGCCAAACCGUUA	1898
PAK3	Sense	
	GAUUAUCGCUGCAAAGGAA	1899
	GAGAGUGCCUGCAAGCUUU	1900
	GACAAGAGGUGGCCAUAAA	1901
	UUAAAUCGCUGUCUUGAGA	1902
PAK4	Sense	
	ACUAAGAGGUGAACAUGUA	1903
	GAUCAUGAAUGUCCGAAGA	1904
	GAUGAGACCCUACUACUGA	1905
	CAGCAAAGGUGCCAAAGAU	1906
PAK6	Sense	
	UAAAGGCAGUUGUCCACUA	1907
	GAAGGGACCUGCUUUCUUG	1908
	GCAAAGACGUCCCUAAGAG	1909
	CCAAUGGGCUGGCUGCAAA	1910
PAK7	Sense	
	GAGCACGGCUUUAUAAGU	1911

	CAAACUCCGUUAUGAUUA	1912
	GGAUAAAGUUGUCUGAUUU	1913
	GGAAAUGCCUCCAUAUA	1914
HDAC1	Sense	
	GGACAUCGCUGUGAAUUGG	1915
	AGAAAGAAGUCACCGAAGA	1916
	GGACAAGGCCACCCAAUGA	1917
	CCACAGCGAUGACUACAUU	1918
HDAC2	Sense	
	GCUGUAAAUAUUGGCUUA	1919
	GCAAAGAAAGCUAGAAUUG	1920
	CAUCAGAGAGUCUUAUAUA	1921
	CCAAUGAGUUGCCAUAUAA	1922
CREBBP	Sense	
	GGCCAUAGCUUAAUUAUUC	1923
	GCACAGCCGUUUACCAUGA	1924
	GGACAGCCCUUUAGUCAAG	1925
	GAACUGAUUCCUGAAAUAA	1926
BTRC	Sense	
	CACAUAAACUCGUAUCUUA	1927
	GAGAAGGCACUCAAGUUUA	1928
	AGACAUAGUUUACAGAGAA	1929
	GCAGAGAGAUUUCAUAACU	1930
RIPK2	Sense	
	GAACAUACCUGUAAAUCAU	1931
	GGACAUCGACCUGUUAUUA	1932
	UAAAUGAACUCCUACAUAG	1933
	GGAUUUAUCUCUGAACAUAA	1934
VAV1	Sense	
	GCAGAAAUACAUCUACUAA	1935
	GCUAUGAGCUGUUCUUA	1936
	CGACAAAGCUCUACUCAUC	1937
	GCUCAACCCUGGAGACAUU	1938
VAV2	Sense	
	GGACAAGACUCGCAGAUUU	1939
	GCUGAGCGCUUUGCAAUAA	1940
	CAAGAAGUCUCACGGGAAA	1941
	UCACAGAGGCCAAGAAAUU	1942
GRB2	Sense	
	UGGAAGCCAUCGCCAAUA	1942
	CAUCAGUGCAUGACGUUUA	1943
	UGAAUGAGCUGGUGGAUUA	1944
	UGCCAAAACUUACCUAUA	1945
PLCG1	Sense	
	GAGCUGCACUCCAUGAGA	1946
	GAAACCAAGCCAUUAUGA	1947
	CCAAGGAGCUACUGACAUU	1948

	AGAGAAACAUGGCCCAAUA	1949
ITGB1	Sense	
	CCACAGACAUUUACAUUAA	1950
	GAAGGGAGUUUGCUAAAUU	1951
	GAACAGAUCUGAUGAAUGA	1952
	CAAGAGAGCUGAAGACUAU	1953
ITGA4	Sense	
	GCAUAUAUAUUCAGCAUUG	1954
	CAACUUGACUGCAGUAUUG	1955
	GAACUUAACUUUCCAUGUU	1956
	GACAAGACCUGUAGUAAUU	1957
STAT1	Sense	
	AGAAAGAGCUUGACAGUAA	1958
	GGAAGUAGUUCACAAAUA	1959
	UGAAGUAUCUGUAUCCAA	1960
	GAGCUUCACUCCCUUAGUU	1961
KRAS2	Sense	
	UAAGGACUCUGAAGAUGUA	1962
	GACAAAGUGUGUAAUUAUG	1963
	GCUCAGGACUUAGCAAGAA	1964
	GAAACUGAAUACCUAAGAU	1965
	GAAACUGAAUACCUAAGAU	1966
	UAAGGACUCUGAAGAUGUA	1967
	GACAAAGUGUGUAAUUAUG	1968
	GCUCAGGACUUAGCAAGAA	1969
HRAS	Sense	
	CCAUCCAGCUGAUCCAGAA	1970
	GAACCCUCCUGAUGAGAGU	1971
	GAGGACAUCCACCAGUACA	1972
BRAF	Sense	
	GAUUAGAGACCAAGGAUUU	
	CCACUGAUGUGUGUAAUU	1973
	CAAUAGAACCUGUCAAUAU	1974
	GAAGACAGGAAUCGAAUGA	1975
ELK1	Sense	
	GAUGUGAGUAGAAGAGUUA	1975
	GGAAGAAUUUGUACCAUUU	1976
	GAACGACCUUUCUUUCUUU	1977
	GGAGUCAUCUCUCCUAUA	1978
RALGDS	Sense	
	GGAGAAGCCUCACCUCUUG	1979
	GCAGAAAGGACUCAAGAUU	1980
	GAGAACAACUACUCAUUGA	1981
	GAACUUCUCGUCACUGUAU	1982
PRKCA	Sense	
	GGAUUGUUCUUUCUUCUAU	1983
	GAAGGGUUCUCGUAUGUCA	1984

MAP2K4	GAAGAAGGAUGUGGUGAUU	1985
	GGACUGGGGAUCGAACAACA	1986
	Sense	
	GGACAGAAGUGGAAAUAUU	1987
	UCAAAGAGGUGAACAUUAA	1988
	GACCAAUCUCAGUUGUUU	1989
MAP2K7	GGAGAAUGGUGCUGUUUAA	1990
	Sense	
	GAAGAGACCAAAGUAUAAU	1991
	GAAGACCGGCCACGUCAUU	1992
	GGAAGAGACCAAAGUAUAA	1993
	GCAUUGAGAUUGACCAGAA	1994
MAPK8	UGAGAGAACGAGAAAGUUG	1995
	GUGAAACCCUGUCUGCAUU	1996
	GGAUCUCUCUCAACAACUA	1997
	ACAACUAGGUGAACACAUA	1998
	Sense	
	UCACAGUCCUGAAACGAUA	1999
MAPK9	GAUUGGAGAUUCUACAUUC	2000
	GCUCAUGGAUGCAAUCUU	2001
	GAAGCUAAGCCGACCAUUU	2002
	Sense	
	AAAGAGAGCUUAUCGUGAA	2003
	GAUGAUAGGUUAGAAAUAG	2004
AIF1	ACAAAGAAGUCAUGGAUUG	2005
	GGAGCUGGAUCAUGAAAGA	2006
	Sense	
	GAAAAGGGAUGAUGGGAUU	2007
	CCUAGACGAUCCCAAUAU	2008
	GAGCCAAACCAGGGAUUUA	2009
BBC3	UGAAACGAAUGCUGGAGAA	2010
	UCACUCACCCAGAGAAUA	2011
	CCAAGAAAGCUAUCUCUGA	2012
	AGACUCACCUAGAGCUAAA	2013
	Sense	
	CCUGGAGGGUCCUGUACAA	2014
BCL2L1	GAGCAAUGAGCCAAACGU	2015
	GGAGGGUCCUGUACAAUCU	2016
	GACUUUCUCUGCACCAUGU	2017
	Sense	
	CCAGGGAGCUUGAAAGUUU	2018
	AAAGUGCAGUUCAGUAAUA	2019
BCL2L11	GAGAAUCACUAACCAGAGA	2020
	GAGCCCAUCCUAUUUAA	2021
	Sense	
	GAGACGAGUUUAACGCUUA	2022
	AAAGCAACCUUCUGAUGUA	2023
	CCGAGAAGGUAGACAAUUG	2024

GCAAAGCAACCUUCUGAUG	2025
AGACAGAGCCACAAGGUAA	2026
GCAAGGAGGUUAGAGAAAU	2027
CAAGGAGGUUAGAGAAUA	2028
UCUUACGACUGUACGUUA	2029

BID	Sense	
	GAAGACAUCAUCCGGAUA	2030
	CAACAGCGUCCUAGAGAA	2031
	GAAAUGGGAUGGACUGAAC	2032
	ACGAUGAGCUGCAGACUGA	2033

BIRC2	Sense	
	GAAAGAAGCCUGCAUAUA	2034
	GAAAUUGACUCUACAUUGU	2035
	ACAAAUAGCACUUAAGGUUA	2036
	GAAUACACCUGUGGUUAAA	2037

BIRC3	Sense	
	GGAGAUGCCUGCCAUUAAA	2038
	UCAUGAUUCUUGUGUUAGA	2039
	GAAAGAACAUGUAAAUGUGU	2040
	GAAGAAAGAACAUGUAAAG	2041

BIRC4	Sense	
	GUAGAUAGAUGGCAUAUG	2042
	GAGGAGGGCUAACUGAUUG	2043
	GAGGAACCCUGCCAUGUAU	2044
	GCACGGAUCUUUACUUUUG	2045

BIRC5	Sense	
	GGCGUAAGAUGAUGGAUUU	2046
	GCAAAGGAAACCAACAAUA	2047
	GCACAAAGCCAUCUAAGU	2048
	CAAAGGAAACCAACAAUA	2049

BRCA1	Sense	
	CCAUACAGCUUCAUAAAUA	2050
	GAAGAGAACUUAUCUAGUG	2051
	GAAGUGGGCUCCAGUAUUA	2052
	GCAAGAUGCUGAUUCAUUA	2053
	CCAUACAGCUUCAUAAAUA	2054

CARD4	Sense	
	GAAAGUUAUGUCAAGGAA	2055
	GAGCAACACUGGCAUAACA	2056
	UACAGAGAUUUGCCUAAA	2057
	GCGAAGAGCUGACCAAAUA	2058

CASP10	Sense	
	CAAAGGGUUCUCUGUUUA	2059
	GAAAUGACCUCUCCUAAGUU	2060
	GAAGGCAGCUGGUUAUAUUC	2061
	GACAUGAUCUCCUUCUGA	2062
	GCACUCUUCUGUCCCUUA	2063

CASP2	Sense	
	GUUUUAAACUCUCCUUUGA	2064
	GCAAGGAGAUGUCUGAAUA	2065
	CAACUCCCCUGAUCUUUAA	2066
	GCUCAAAGAUGUAAUGUAG	2067
CDKN1A	Sense	
	GAACAAGGAGUCAGACAUU	2068
	AAACUAGGCGGUUGAAUGA	2069
	GAUGGAACUUCGACUUUGU	2070
	GUAAACAGAUGGCACUUUG	2071
CFLAR	Sense	
	GAUGUGUCCUCAUUAAUUU	2072
	GAAGAGAGAUACAAGAUGA	2073
	GAGCAUACCUGAAGAGAGA	2074
	GCUAUGAAGUCCAGAAAUU	2075
CLK2	Sense	
	GUGAAUAUGUGAAAUAGUG	2076
	AAAGCAUGCUAGAGUAUGA	2077
	UUAAGAAUGUGGAGAAGUA	2078
	GAUAACAAGCUGACACAUU	2079
CLSPN	Sense	
	GGACGUAUUUGAUGAAGUA	2080
	GCAGAUGGGUUCUAAAUG	2081
	CAAAUGAGGUUGAGGAAAU	2082
	GGAAAUACCUGGAGGAUGA	2083
CSNK2A 1	Sense	
	GAUCCACGUUUCAAUGAUA	2084
	GCAUUUAGGUGGAGACUUC	2085
	GAUGUACGAUUUAGUUUG	2086
	UGAAUUAGAUAUCCACGUUUC	2087
CTNNB1	Sense	
	GCACAAGAAUGGAUCACAA	2088
	GCUGAAACAUGCAGUUGUA	2089
	GUACGUACCAUGCAGAAUA	2090
	GAACUUGCAUUGUGAUUGG	2091
CXCR4	Sense	
	GAAGCAUGACGGACAAGUA	2092
	GAACAUUCCAGAGCGUGUA	2093
	GUUCUUAGUUGCUGUAUGU	2094
	CAUCAUGGUUGGCCUUAUC	2095
CXCR6	Sense	
	GGAACAAACUGGCAAAGCA	2096
	GAUCAGAGCAGCAGUGAAA	2097
	GGGCAAAACUGAAUUUAUA	2098
	GAUCUCAGGUUCUCCUUGA	2099
DAXX	Sense	

	CUACAGAUCUCCAAUGAAA	2100
	GCUACAAGCUGGAGAAUGA	2101
	GGAAACAGCUAUGUGGAAA	2102
	GGAGUUGGAUCUCUCAGAA	2103
GAS41	Sense	
	GUAGUAAGCUAAACUGAAA	2104
	GACAAUAUGUUCAAGAGAA	2105
	GACAACAUCUCGUCAGCUA	2106
	UAUAUGAUGUGUCCAGUAA	2107
GTSE1	Sense	
	CAAAGAAGCUCACUUCUG	2108
	GAACAGCCCUAAAGUGGUU	2109
	GAACAUGGAUGACCCUAG	2110
	GGGCAAAGCUAAAUCAAGU	2111
HDAC3	Sense	
	GGAAAGCGAUGUGGAGAUU	2112
	CCAAGACCGUGGCCUAUUU	2113
	AAAGCGAUGUGGAGAUUUA	2114
	GUGAGGAGCUUCCCUAUAG	2115
HDAC5	Sense	
	GAAUCCUCUUGUCGAAGU	2116
	GUUAUUAGCACCUUUAAGA	2117
	GGAGGGAGGCCAUGACUUG	2118
	CAGGAGAGCUCAAGAAUGG	2119
	GGAUAUGGAUUUCAGUAA	2120
	GGAAGUCGGUGCCUUGGUU	2121
	GGAAGGAGAGACUGGUUU	2122
HEC	Sense	
	GCAGAUACUUGCACGGUUU	2123
	GAGUAGAACUAGAAUGUGA	2124
	GCGAAUAAAUCAUGAAAGA	2125
	GAAGAUGGAUUUAUGCAUA	2126
HIST1H2 AA	Sense	
	GGCAAUGCGUCUCGCGAUA	2127
	GAUCCGCAAUGAUGAGGAA	2128
	GCAAUGCGUCUCGCGAUAA	2129
	GAGGAACUCAAUAAAGCUUU	2130
LMNB1	Sense	
	AAUAGAAGCUGUGCAAUUA	2131
	CAACUGACCUCAUCUGGAA	2132
	GAAGGAAUCUGAUCUUAUU	2133
	GGGAAGGGUUUCUCUAUUA	2134
LMNB2	Sense	
	GGAGGUUCAUUGAGAAUUG	2134
	GGCAAUAGCUCACCGUUUA	2135
	CAAAUACGCUUAGCUGUGU	2136
	GGAGAUCGCCUACAAGUUC	2137

MYB	Sense	
	GCAGAAACACUCCAAUUUA	2138
	GUAAAUACGUGAAUGCAUU	2139
	GCACUGAACUUUUGAGAU	2140
	GAAGAACAGUCAUUUGAUG	2141
MYT1	Sense	
	GAGGUGAGCUGUAAAUA	2142
	GCAGGUGAUUCCUAAUA	2143
	GGGAGAAGAUUUUAAUUG	2144
	CAACUUCUCUCCUGAACUU	2145
NFKBIB	Sense	
	GGACACGGCACUGCACUUG	2146
	GCACUUGGCUGUGAUUCAU	2148
	GAGACGAGGGCGAUGAAUA	2149
	CAUGAACCCUCCUGGAUU	2150
NFKBIA	Sense	
	GAACAUGGACUUGUAUUAU	2151
	GAUGUGGGGUGAAAAGUUA	2152
	GGACGAGAAAGAUCAUUGA	2153
	AGGACGAGCUGCCCUAUGA	2154
NFKBIE	Sense	
	GAAGGGAAGUUUCAGUAAC	2155
	GGAAGGGAAGUUUCAGUAA	2156
	GGAAACUGCUGCUGUGUAC	2157
	GAACCAACCACUCAUGGAA	2158
NUMA1	Sense	
	GGAACAGUUUGAAUAUAA	2159
	GCAGUAGCEUGAAGCAGAA	2160
	CGAGAAGGAUGCACAGAU	2161
	GCAAGAGGCUGAGAGGAAA	2162
NUP153	Sense	
	GAAGACAAAUGAAAGCUAA	2163
	GAUAAAGACUGCUGUUAGA	2164
	GAGGAGAGCUCUAAUAUUA	2165
	GAGGAAGCCUGAUUAAAGA	2166
OPA1	Sense	
	GAAAGAGCAUGAUGACAU	2167
	GAGGAGAGCUCUAUUAUGU	2168
	GAAACUGAAUGGAAGAAUA	2169
	AAAGAAGGCUGUACCGUUA	2170
PARVA	Sense	
	CUACAUGUCUUUGCUCUUA	2171
	GCUAAGUCCUGUAAGAAUA	2172
	CAAAGGCAAUGUACUGUUU	2173
	GAACAAUGGUGGAUCCAAA	2174
PIK3CG	Sense	
	AAGUUCAGCUUCUCUAUUA	2175

	GAAGAAUCUCUGAUGGAU	2176
	GAACACCUUACUCUAUAA	2177
	GCAUGGAGCUGGAGAACUA	2178
PRKDC	Sense	
	GAUGAAAGCUCUAAAGAUG	2179
	GAAAGGAGGUUCUAAACUA	2180
	GGAAGAAGCUCAUUUGAUU	2181
	GCAAAGAGGUGGCAGUUA	2182
RASA1	Sense	
	GGAAGAAGAUCCACAUGAA	2183
	GAACAUACUUUCAGAGCUU	2184
	GAACAAUCUUUUGCUGUAUA	2185
	UAACAGAACUGCUUCAACA	2186
SLC9A1	Sense	
	GAAGAGAUCCACACACAGU	2187
	UCA AUGAGCUGCUGCACAU	2188
	GAAGAUAGGUUCCAUGUG	2189
	GAAUUACCCUCCUCAUCU	2190
TEGT	Sense	
	CUACAGAGCUUCAGUGUGA	2191
	GAACAUUUUGAUCGAAAG	2192
	GAGCAAACCUAGAUAAAGGA	2193
	GCAUUGAUCUCUUCUAGA	2194
TERT	Sense	
	GGAAGACAGUGGUGAACUU	2195
	GCAAAGCAUUGGAAUCAGA	2196
	GAGCUGACGUGGAAGAUGA	2197
	GAACGGGCCUGGAACCAUA	2198
TNFRSF 6	Sense	
	GAUACUAACUGCUCUCAGA	2199
	GAAAGAAUGGUGUCAUGA	2200
	UCAUAAUGUCCCAUGUAA	2201
	UCAUGAAUCUCCAACCUUA	2202
	GAUGUUGACUUGAGUAAAU	2203
TOP1	Sense	
	GAAAGGAAAUGACUAAUGA	2204
	GAAGAAGGCUGUUCAGAGA	2205
	GGAAGUAGCUACGUUCUUU	2206
	GGACAUAAAGUGGAAAGAAG	2207
TOP2A	Sense	
	GAAAGAGUCCAUCAGAUUU	2208
	CAAACUACAUUGGCAUUUA	2209
	AAACAGACAUGGAUGGAUA	2210
	CGAAAGGAAUGGUUAACUA	2211
TOP3A	Sense	
	CCAGAAAUCUCCACAGAA	2212

GAAACUAUCUGGAUGUGUA	2213
CCACAAAGAUGGUAUCGUA	2214
GGAAAUGGCUGUGGUAACA	2215

TOP3B

Sense	
GAGACAAGAUGAAGACUGU	2216
GCACAUGGGCUGCGUCUUU	2217
CCAGUGCGCUUCAAGAUGA	2218
GAACAUCUGCUUUGAGGUU	2219

WEE1

Sense	
GGUAUUGCCUUGUGAAUUU	2220
GCAGAACAAUACGAAUAG	2221
GUACAUAGCUGUUUGAAAU	2222
GCUGUAAACUUGUAGCAU	2223

In addition, to identifying functional siRNA against gene families or pathways, it is possible to design duplexes against genes known to be involved in specific diseases. For example when dealing with human disorders associated with allergies, it will be beneficial to develop siRNA against a number of genes including but not limited to:

the interleukin 4 receptor gene

(SEQ. ID NO. 2224: UAGAGGUGCUCAUUCAUUU,

10 SEQ. ID NO. 2225: GGUAUAAGCCUUUCCAAGA,
SEQ. ID NO. 2225: ACACACAGCUGGAAGAAAU,
SEQ. ID NO. 2226: UAACAGAGCUUCCUUAGGU),

the Beta-arrestin-2

15 (SEQ. ID NO. 2227: GGAUGAAGGAUGACGACUA,
SEQ. ID NO. 2228: ACACCAACCUCAUUGAAUU,
SEQ. ID NO. 2229: CGAACAAGAUGACCAGGUA,
SEQ. ID NO. 2230: GAUGAAGGAUGACGACUAU,),

20 the interferon-gamma receptor 1 gene

(SEQ. ID NO. 2231: CAGCAUGGCUCUCCUCUUU,
SEQ. ID NO. 2232: GUAAAGAACUAUGGUGUUA,
SEQ. ID NO. 2233: GAAACUACCUGUUACAUUA,
SEQ. ID NO. 2234: GAAGUGAGAUCAGUAUAA),

25

the matrix metalloproteinase MMP-9

(SEQ. ID NO. 2235: GGAACCAGCUGUAUUUGUU,
SEQ. ID NO. 2236: GUUGGAGUGUUUCUAAUAA,
SEQ. ID NO. 2237: GCGCUGGGCUUAGAUCU,
5 SEQ. ID NO. 2238: GGAGCCAGUUUGCCGGAUA),

the Slc11a1 (Nramp1) gene

(SEQ. ID NO. 2239: CCAAUGGCCUGCUGAACAA,
SEQ. ID NO. 2240: GGGCCUGGCUUCCUCAUGA,
10 SEQ. ID NO. 2241: GGGCAGAGCUCCACCAUGA,
SEQ. ID NO. 2242: GCACGGCCAUUGCAUUCAA),

SPINK5

(SEQ. ID NO. 2243: CCAACUGCCUGUCAAUAA,
15 SEQ. ID NO. 2244: GGAUACAUGUGAUGAGUUU,
SEQ. ID NO. 2245: GGACGAAUGUGCUGAGUAU,
SEQ. ID NO. 2246: GAGCUUGUCUUAUUUGCUA),

the CYP1A2 gene

20 (SEQ. ID NO. 2247: GAAAUGCUGUGUCUUCGUA,
SEQ. ID NO. 2248: GGACAGCACUUCCCUGAGA,
SEQ. ID NO. 2249: GAAGACACCACCAUUCUGA,
SEQ. ID NO. 2250: GGCCAGAGCUUGACCUUCA),

25 thymosin-beta4Y

(SEQ. ID NO. 2251: GGACAGGCCUGCGUUGUUU,
SEQ. ID NO. 2252: GGAAAGAGGAAGCUCAUGA,
SEQ. ID NO. 2253: GCAAACACGUUGGAUGAGU,
SEQ. ID NO. 2254: GGACUAUGCUGCCCUIIUUG,

30

activin A receptor IB

(SEQ. ID NO. 2255: ACAAGACGCUCCAGGAUCU,
SEQ. ID NO. 2254: GCAACAGGAUCGACUUGAG,
SEQ. ID NO. 2255: GAAGCUGCGUCCCAACAUC,

SEQ. ID NO. 2256: GCAUAGGCCUGUAAUCGUA,
SEQ. ID NO. 2257: UCAGAGAGUUCGAGACAAA,
SEQ. ID NO. 2258: UGCGAAAGGUUGUAUGUGA,
SEQ. ID NO. 2259: GCAACAGGAUCGACUUGAG,
5 SEQ. ID NO. 2260: GAAUAGCGUUGUGUGUUAU,
SEQ. ID NO. 2261: UGAAUAGCGUUGUGUGUUA,
SEQ. ID NO. 2262: GGGAUUCAGUUUGUUGAAUA,
SEQ. ID NO. 2263: GAGCCUGAAUCAUCGUUUA,),

10 ADAM33

(SEQ. ID NO. 2264: GGAAGUACCUGGAACUGUA,
SEQ. ID NO. 2265: GGACAGAGGGAACCAUUUA,
SEQ. ID NO. 2266: GGUGAGAGGUAGCUCCUAA,
SEQ. ID NO. 2267: AAAGACAGGUGGCCACUGA),

15

the TAP1 gene

(SEQ. ID NO. 2268: GAAAGAUGAUCAGCUAUUU,
SEQ. ID NO. 2269: CAACAGAACCAGACAGGUA,
SEQ. ID NO. 2270: UGAGAAAUGUUCAGAAUGU,
20 SEQ. ID NO. 2271: UACCUUCACUCGAAACUUA,

COX-2

(SEQ. ID NO. 2272: GAACGAAAGUAAAGAUGUU,
SEQ. ID NO. 2273: GGACUUAUGGGUAAUGUUA,
25 SEQ. ID NO. 2274: UGAAAGGACUUAUGGGUAA,
SEQ. ID NO. 2275: GAUCAGAGUUCACUUUCUU),

ADPRT

(SEQ. ID NO. 2276: GGAAAGAUGUUAAGCAUUU,
30 SEQ. ID NO. 2277: CAUGGGAGCUCUUGAAAUA,
SEQ. ID NO. 2278: GAACAAGGAUGAAGUGAAG,
SEQ. ID NO. 2279: UGAAGAAGCUCACAGUAAA,),

HDC

(SEQ. ID NO. 2280: CAGCAGACCUUCAGUGUGA,
SEQ. ID NO. 2281: GGAGAGAGAUGGUGGAUUA,
SEQ. ID NO. 2282: GUACAGAGCUGGAGAUGAA,
SEQ. ID NO. 2283: GAACGUCCCUUCAGUCUGU),

5

HnmT

(SEQ. ID NO. 2284: CAAAUUCUCUCCAAAGUUC,
SEQ. ID NO. 2285: GGAUAUAUCUGACUGCUUU,
SEQ. ID NO. 2286: GAGCAGAGCUUGGGAAAGA,
10 SEQ. ID NO. 2287: GAUAUGAGAUGUAGCAAAU),

GATA-3

(SEQ. ID NO. 2288: GAACUGCUUUCUUUCGUUU,
SEQ. ID NO. 2289: GCAGUAUCAUGAAGCCUAA,
15 SEQ. ID NO. 2290: GAAACUAGGUCUGAUAUUC,
SEQ. ID NO. 2291: GUACAGCUCCGGACUCUUC),

Gab2

(SEQ. ID NO. 2292: GCACAACCAUUCUGAAGUU,
20 SEQ. ID NO. 2293: GGACUUAGAUGCCAGGAUG,
SEQ. ID NO. 2294: GAAGGUGGAUUCUAGGAAA,
SEQ. ID NO. 2295: GGACUAGCCCUGCUGUUUA), and

STAT6

(SEQ. ID NO. 2296: GAUAGAAACUCCUGCUAAU,
25 SEQ. ID NO. 2297: GGACAUUUUAUCCCCAGCUA,
SEQ. ID NO. 2298: GGACAGAGCUACAGACCUA,
SEQ. ID NO. 2299: GGAUGGCUCUCCACAGAU).

30 In addition, rationally designed siRNA or siRNA pools can be directed against
genes involved in anemia, hemophila or hypercholesterolemia. Such genes would
include, but are not be limited to:

APOA5

(SEQ. ID NO. 2300: GAAAGACAGCCUUGAGCAA,

SEQ. ID NO. 2301: GGACAGGGAGGCCACCAAA,
SEQ. ID NO. 2302: GGACGAGGCUUGGGCUUUG,
SEQ. ID NO. 2303: AGCAAGACCUCAACAAUUAU),

5 HMG-CoA reductase

(SEQ. ID NO. 2304: GAAUGAAGCUUUGCCCUUU,
SEQ. ID NO. 2305: GAACACAGUUUAGUGCUUU,
SEQ. ID NO. 2306: UAUCAAGAGCUCUUA AUGUU,
SEQ. ID NO. 2307: UGAAGAAUGUCUACAGAU),

10

NOS3

(SEQ. ID NO. 2308: UGAAGCACCUGGAGAAUGA,
SEQ. ID NO. 2309: CGGAACAGCACAAGAGUUA,
SEQ. ID NO. 2310: GGAAGAAGACCUUUAAGA,

15 SEQ. ID NO. 2309: GCACAAGAGUUAUAAGAUC),

ARH

(SEQ. ID NO. 2310: CGAUACAGCUUGGCACUUU,
SEQ. ID NO. 2311: GAGAAGCGCUGCCCUGUGA,
SEQ. ID NO. 2312: GAAUCAUGCUGUUCUUCUUU,
SEQ. ID NO. 2313: GGAGUAACCGGACACCUUA),

20

CYP7A1

(SEQ. ID NO. 2314: UAAGGUGACUCGAGUGUUU,
SEQ. ID NO. 2315: AAACGACACUUUCAUCAA,
SEQ. ID NO. 2316: GGACUCAAGUUAAGUAUU,
SEQ. ID NO. 2317: GUAAUGGACUCAAGUAAA),

25

FANCA

(SEQ. ID NO. 2318: GGACAUCACUGCCCACUUC,
SEQ. ID NO. 2319: AGAGGAAGAUGUUCACUUA,
SEQ. ID NO. 2320: GAUCGUGGCUCUUCAGGAA,
SEQ. ID NO. 2321: GGACAGAGGCAGAUAAAGAA),

30

FANCG

(SEQ. ID NO. 2322: GCACUAAGCAGCCUUCAUG,
SEQ. ID NO. 2323: GCAAGCAGGUGCCUACAGA,
SEQ. ID NO. 2324: GGAAUUAGAUGCUCCAUUG,
5 SEQ. ID NO. 2325: GGACAUCUCUGCCAAAGUC),

ALAS

(SEQ. ID NO. 2326: CAAUAUGCCUGGAAACUAU,
SEQ. ID NO. 2327: GGUUAAGACUCACCAGUUC,
10 SEQ. ID NO. 2328: CAACAGGACUUUAGGUUCA,
SEQ. ID NO. 2329: GCAUAAGAUUGACAUCAUC),

PIGA

(SEQ. ID NO. 2330: GAAAGAGGGCAUAAGGUUA,
15 SEQ. ID NO. 2331: GGACUGAUCUUUAAACUAU,
SEQ. ID NO. 2332: UCAAAUGGCUUACUUCAUC,
SEQ. ID NO. 2333: UCUAAGAACUGAUGUCUAA), and

factor VIII

20 (SEQ. ID NO. 2334: GCAAAUAGAUCUCCAUAUAC,
SEQ. ID NO. 2335: CCAGAUUAUGUCGUUCUUUA,
SEQ. ID NO. 2336: GAAAGGCUGUGCUCUCAA,
SEQ. ID NO. 2337: GGAGAAACCUGCAUGAAAG,
SEQ. ID NO. 2338: CUUGAAGCCUCCUGAAUUA,
25 SEQ. ID NO. 2339: GAGGAAGCAUCCAAAGAUU,
SEQ. ID NO. 2340: GAUAGGAGAUACAAACUUU).

Furthermore, rationally designed siRNA or siRNA pools can be directed
against genes involved in disorders of the brain and nervous system. Such genes
30 would include, but are not be limited to:

APBB1

(SEQ. ID NO. 2341: CUACGUAGCUCGUGAUAAG,
SEQ. ID NO. 2342: GCAGAGAUGUCCACACGUU,
SEQ. ID NO. 2343: CAUGAGAUCUGCUCUAAGA,

SEQ. ID NO. 2344: GGGCACCUCUGCUGUAUUG),

BACE1

(SEQ. ID NO. 2345: CCACAGAGCAAGUGAUUUA,

5 SEQ. ID NO. 2346: GCAGAAAGGAGAUCAUUUA,

SEQ. ID NO. 2347: GUAGCAAGAUCUUUACAUA,

SEQ. ID NO. 2348: UGUCAGAGCUUGAUUAGAA),

PSEN1

10 (SEQ. ID NO. 2349: GAGCUGACAUUGAAAUAUG,

SEQ. ID NO. 2350: GUACAGCUAUUUCUCAUCA,

SEQ. ID NO. 2351: GAGGUUAGGUGAAGUGGUU,

SEQ. ID NO. 2352: GAAAGGGAGUCACAAGACA,

SEQ. ID NO. 2353: GAACUGGAGUGGAGUAGGA,

15 SEQ. ID NO. 2354: CAGCAGGCAUAUCUCAUUA,

SEQ. ID NO. 2355: UCAAGUACCUCUCCUGAAUG),

PSEN2

(SEQ. ID NO. 2356: GCUGGGAAGUGGCUUAAUA,

20 SEQ. ID NO. 2357: CAUAUUCCCUGCCCUGAUA,

SEQ. ID NO. 2358: GGGAAGUGCUCAAGACCUA,

SEQ. ID NO. 2359: CAUAGAAAGUGACGUGUUA),

MASS1

25 (SEQ. ID NO. 2360: GGAAGGAGCUGUUAUGAGA,

SEQ. ID NO. 2361: GAAAGGAGAAGCUAAAUUA,

SEQ. ID NO. 2362: GGAGGAAGGUCAAGAUUUA,

SEQ. ID NO. 2363: GGAAAUAGCUGAGAUAAUG,),

30 ARX

(SEQ. ID NO. 2364: CCAGACGCCUGAUUAUUGAA,

SEQ. ID NO. 2365: CAGCACCACUCAAGACCAA,

SEQ. ID NO. 2366: CGCCUGAUUAUUGAAGUAAA,

SEQ. ID NO. 2367: CAACAUCCACUCUCUCUUG) and

NNMT

(SEQ. ID NO. 2368: GGGCAGUGCUCCAGUGGUA,

SEQ. ID NO. 2369: GAAAGAGGCUGGCUACACA,

5 SEQ. ID NO. 2370: GUACAGAAGUGAGACAUAA,

SEQ. ID NO. 2371: GAGGUGAUCUCGCAAAGUU).

In addition, rationally designed siRNA or siRNA pools can be directed against genes involved in hypertension and related disorders. Such genes would include, but

10 are not be limited to:

angiotensin II type 1 receptor

(SEQ. ID NO. 2372: CAAGAAGCCUGCACCAUGU,

SEQ. ID NO. 2373: GCACUUCACUACCAAUAUGA,

SEQ. ID NO. 2374: GCACUGGUCCCAAGUAGUA,

15 SEQ. ID NO. 2375: CCAAAGGGCAGUAAAGUUU,

SEQ. ID NO. 2376: GCUCAGAGGAGGUGUAUUU,

SEQ. ID NO. 2377: GCACUUCACUACCAAUAUGA,

SEQ. ID NO. 2378: AAAGGGCAGUAAAGUUU),

20 AGTR2

(SEQ. ID NO. 2379: GAACAUCUCUGGCAACAAU,

SEQ. ID NO. 2380: GGUGAUUAUAUCUCAAUUG,

SEQ. ID NO. 2381: GCAAGCAUCUUAUAUAGUU,

SEQ. ID NO. 2382: GAACCAGUCUUUCAACUCA), and other related targets.

25

Example XIII: Validation of Multigene Knockout using Rab5 and Eps

Two or more genes having similar, overlapping functions often leads to

genetic redundancy. Mutations that knockout only one of, *e.g.*, a pair of such genes

(also referred to as homologs) results in little or no phenotype due to the fact that the

30 remaining intact gene is capable of fulfilling the role of the disrupted counterpart. To

fully understand the function of such genes in cellular physiology, it is often

necessary to knockout or knockdown both homologs simultaneously. Unfortunately,

concomitant knockdown of two or more genes is frequently difficult to achieve in

higher organisms (*e.g.* mice) thus it is necessary to introduce new technologies dissect

gene function. One such approach to knocking down multiple genes simultaneously is by using siRNA. For example, **Figure 11** showed that rationally designed siRNA directed against a number of genes involved in the clathrin-mediated endocytosis pathway resulted in significant levels of protein reduction (e.g. >80%). To determine the effects of gene knockdown on clathrin-related endocytosis, internalization assays were performed using epidermal growth factor and transferrin. Specifically, mouse receptor-grade EGF (Collaborative Research Inc.) and iron-saturated human transferrin (Sigma) were iodinated as described previously (Jiang, X., Huang, F., Marusyk, A. & Sorkin, A. (2003) *Mol Biol Cell* 14, 858-70). HeLa cells grown in 12-well dishes were incubated with ^{125}I -EGF (1 ng/ml) or ^{125}I -transferrin (1 $\mu\text{g/ml}$) in binding medium (DMEM, 0.1% bovine serum albumin) at 37°C, and the ratio of internalized and surface radioactivity was determined during 5-min time course to calculate specific internalization rate constant k_e as described previously (Jiang, X *et al.*). The measurements of the uptakes of radiolabeled transferrin and EGF were performed using short time-course assays to avoid influence of the recycling on the uptake kinetics, and using low ligand concentration to avoid saturation of the clathrin-dependent pathway (for EGF Lund, K. A., Opresko, L. K., Starbuck, C., Walsh, B. J. & Wiley, H. S. (1990) *J. Biol. Chem.* 265, 15713-13723).

The effects of knocking down Rab5a, 5b, 5c, Eps, or Eps 15R (individually) are shown in **Figure 22** and demonstrate that disruption of single genes has little or no effect on EGF or Tfn internalization. In contrast, simultaneous knock down of Rab5a, 5b, and 5c, or Eps and Eps 15R, leads to a distinct phenotype (note: total concentration of siRNA in these experiments remained constant with that in experiments in which a single siRNA was introduced, see **Figure 23**). These experiments demonstrate the effectiveness of using rationally designed siRNA to knockdown multiple genes and validates the utility of these reagents to override genetic redundancy.

Example XIV. Validation of Multigene Targeting Using G6PD, GAPDH, PLK, and UQC.

Further demonstration of the ability to knock down expression of multiple genes using rationally designed siRNA was performed using pools of siRNA directed

against four separate genes. To achieve this, siRNA were transfected into cells (total siRNA concentration of 100nM) and assayed twenty-four hours later by B-DNA. Results shown in **Figure 24** show that pools of rationally designed molecules are capable of simultaneously silencing four different genes.

5

Example XV. Validation of Multigene Knockouts As Demonstrated by Gene Expression Profiling, a Prophetic Example

To further demonstrate the ability to concomitantly knockdown the expression of multiple gene targets, single siRNA or siRNA pools directed against a collection of genes (*e.g.* 4, 8, 16, or 23 different targets) are simultaneously transfected into cells and cultured for twenty-four hours. Subsequently, mRNA is harvested from treated (and untreated) cells and labeled with one of two fluorescent probes dyes (*e.g.* a red fluorescent probe for the treated cells, a green fluorescent probe for the control cells.). Equivalent amounts of labeled RNA from each sample is then mixed together and hybridized to sequences that have been linked to a solid support (*e.g.* a slide, "DNA CHIP"). Following hybridization, the slides are washed and analyzed to assess changes in the levels of target genes induced by siRNA.

Example XVI. Identifying Hyperfunctional siRNA

20

Identification of Hyperfunctional Bcl-2 siRNA

The ten rationally designed Bcl2 siRNA (identified in **Figure 13, 14**) were tested to identify hyperpotent reagents. To accomplish this, each of the ten Bcl-2 siRNA were individually transfected into cells at a 300pM (0.3nM) concentrations. Twenty-four hours later, transcript levels were assessed by B-DNA assays and compared with relevant controls. As shown in **Figure 25**, while the majority of Bcl-2 siRNA failed to induce functional levels of silencing at this concentration, siRNA 1 and 8 induced >80% silencing, and siRNA 6 exhibited greater than 90% silencing at this subnanomolar concentration.

30

By way of prophetic examples, similar assays could be performed with any of the groups of rationally designed genes described in Example VII or Example VIII. Thus for instance, rationally designed siRNA sequences directed against PDGFA

(SEQ. ID NO. 2383: GGUAAGAUAUUGUGCUUUA,
SEQ. ID NO. 2384: CCGCAAAUAUGCAGAAUUA,
SEQ. ID NO. 2385: GGAUGUACAUGGCGUGUUA,
SEQ. ID NO. 2386: GGUGAAGUUUGUAUGUUUA), or

5

PDGFB

(SEQ. ID NO. 2387: GCUCCGCGCUUCCGAUUU,
SEQ. ID NO. 2388: GAGCAGGAAUGGUGAGAUG,
SEQ. ID NO. 2389: GAACUUGGGAUAAGAGUGU,
10 SEQ. ID NO. 2390: CCGAGGAGCUUUAUGAGAU,
SEQ. ID NO. 2391: UUUAUGAGAUGCUGAGUGA)

could be introduced into cells at increasingly limiting concentrations to determine whether any of the duplexes are hyperfunctional. Similarly, rationally designed sequences directed against

15 HIF1 alpha...

(SEQ. ID NO. 2392: GAAGGAACCUGAUGCUUUA,
SEQ. ID NO. 2393: GCAUAUAUCUAGAAGGUAU,
SEQ. ID NO. 2394: GAACAAAUACAUGGGAUUA,
SEQ. ID NO. 2395: GGACACAGAUUUAGACUUG), or

20

VEGF

(SEQ. ID NO. 2396: GAACGUACUUGCAGAUGUG,
SEQ. ID NO. 2397: GAGAAAGCAUUUGUUUGUA,
SEQ. ID NO. 2398: GGAGAAAGCAUUUGUUUGU,
25 SEQ. ID NO. 2399: CGAGGCAGCUUGAGUUAAA) could be introduced into cells at increasingly limiting concentrations and screened for hyperfunctional duplexes.

Example XVII: Gene Silencing: Prophetic Example

Below is an example of how one might transfect a cell.

- 30 a. Select a cell line. The selection of a cell line is usually determined by the desired application. The most important feature to RNAi is the level of expression of the gene of interest. It is highly recommended to use cell lines for which siRNA transfection conditions have been specified and validated.

- 5 b. Plate the cells. Approximately 24 hours prior to transfection, plate the cells at the appropriate density so that they will be approximately 70 – 90% confluent, or approximately 1×10^5 cells/ml at the time of transfection. Cell densities that are too low may lead to toxicity due to excess exposure and uptake of transfection reagent-siRNA complexes. Cell densities that are too high may lead to low transfection efficiencies and little or no silencing. Incubate the cells overnight. Standard incubation conditions for mammalian cells are 37°C in 5% CO₂. Other cell types, such as insect cells, require different temperatures and CO₂ concentrations that are readily ascertainable by persons skilled in the art. Use conditions appropriate for the cell type of interest.
- 10
- c. SiRNA re-suspension. Add 20 µl siRNA universal buffer to each siRNA to generate a final concentration of 50 µM.
- 15 d. SiRNA-lipid complex formation. Use RNase-free solutions and tubes. Using the following table, Table XI:

e.

Table XI		
	96-well	24-well
Mixture 1 (TransIT-TKO-Plasmid dilution mixture)		
Opti-MEM	9.3 μ l	46.5 μ l
TransIT-TKO (1 μ g/ μ l)	0.5 μ l	2.5 μ l
Mixture 1 Final Volume	10.0 ml	50.0 μl
Mixture 2 (siRNA dilution mixture)		
Opti-MEM	9.0 μ l	45.0 μ l
siRNA (1 μ M)	1.0 μ l	5.0 μ l
Mixture 2 Final Volume	10.0 ml	50.0 μl
Mixture 3 (siRNA-Transfection reagent mixture)		
Mixture 1	10 μ l	50 μ l
Mixture 2	10 μ l	50 μ l
Mixture 3 Final Volume	20 μl	100 μl
Incubate 20 minutes at room temperature.		
Mixture 4 (Media-siRNA/Transfection reagent mixture)		
Mixture 3	20 μ l	100 μ l
Complete media	80 μ l	400 μ l
Mixture 4 Final Volume	100 μl	500 μl
Incubate 48 hours at 37°C.		

- 5 Transfection. Create a Mixture 1 by combining the specified amounts of OPTI-MEM serum free media and transfection reagent in a sterile polystyrene tube. Create a Mixture 2 by combining specified amounts of each siRNA with OPTI-MEM media in sterile 1 ml tubes. Create a Mixture 3 by combining specified amounts of Mixture 1 and Mixture 2. Mix gently (do not vortex) and incubate at room temperature for 20
- 10 minutes. Create a Mixture 4 by combining specified amounts of Mixture 3 to complete media. Add appropriate volume to each cell culture well. Incubate cells with transfection reagent mixture for 24 – 72 hours at 37°C. This incubation time is flexible. The ratio of silencing will remain consistent at any point in the time period. Assay for gene silencing using an appropriate detection method such as RT-PCR,
- 15 Western blot analysis, immunohistochemistry, phenotypic analysis, mass

spectrometry, fluorescence, radioactive decay, or any other method that is now known or that comes to be known to persons skilled in the art and that from reading this disclosure would be useful with the present invention. The optimal window for observing a knockdown phenotype is related to the mRNA turnover of the gene of interest, although 24 – 72 hours is standard. Final Volume reflects amount needed in each well for the desired cell culture format. When adjusting volumes for a Stock Mix, an additional 10% should be used to accommodate variability in pipetting, *etc.* Duplicate or triplicate assays should be carried out when possible.

10 While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departure from the present disclosure as come within known or customary practice
15 within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

20

Claims

1. A method for selecting siRNA comprising selecting an siRNA molecule of 19 – 25 nucleoside bases, said method comprising:
- (a) selecting a target gene;
- 5 (b) measuring the functionality of sequences of 19 – 25 nucleotides in length that are substantially complementary to a stretch of nucleotides of the target sequence, wherein said functionality is dependent upon non-target specific criteria.
- 10 2. The method according to claim 1 wherein said functionality is determined by applying one of the following formulas:

Formula I = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

Formula II = $-(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$;

Formula III = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3$;

Formula IV = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

Formula V = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

Formula VI = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$;

Formula VII = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2$;

wherein in Formulas I – VII:

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense

strand at positions 15 –19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_{17}-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$

wherein

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
5 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
10 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
15 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
20 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

25 $G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
30 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;

- $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
5 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;
10
 GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand
or within positions 15 –18 if the sense strand is only 18 base pairs in length;
 GC_{total} = the number of G and C bases in the sense strand;
 $T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs,
15 otherwise its value is 0; and
 X = the number of times that the same nucleotide repeats four or more times in a
row.
- 20 3. A method of gene-silencing comprising selecting an siRNA according to
claim 2 and introducing it into a cell.
 4. The method according to claim 3 wherein said introducing is by allowing
passive uptake of the siRNA.
 - 25 5. The method according to claim 3, wherein said introducing is through the
use of a vector.
 - 30 6. A method for developing an siRNA algorithm for selecting siRNA, said
method comprising:
 - (a) selecting a set of siRNA;
 - (b) measuring the gene silencing ability of each siRNA from said set;
 - (c) determining the relative functionality of each siRNA;
 - (d) determining the amount of improved functionality by the presence or
absence of at least one variable selected from the group consisting of

- the total GC content, melting temperature of the siRNA, GC content at positions 15 –19, the presence or absence of a particular nucleotide at a particular position and the number of times that the same nucleotide repeats within a given sequence; and
- 5 (e) developing an algorithm using the information of step (d).
7. A method of selecting an siRNA with improved functionality, said method comprising using the algorithm of claim 6.
- 10 8. A method of selecting hyperfunctional siRNA, said method comprising using at least one functional siRNA, wherein at least one said functional siRNA has been selected according to the method of claim 7 and measuring the silencing ability of said at least one functional siRNA, wherein silencing ability is measured at a concentration of less than 1
- 15 nanomolar siRNA.
9. An siRNA molecule, wherein said siRNA molecule is effective at silencing Bcl-2.
- 20 10. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a sequence substantially similar to a sequence selected from the group consisting of GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301); GAAGUACAUCCAUUAUAAG (SEQ. ID NO. 302); GUACGACAACCGGAGAUUA (SEQ. ID NO. 303);
- 25 AGAUAGUGAUGAAGUACAU (SEQ. ID NO. 304); UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305); CAUGCGGCCUCUGUUUGA (SEQ. ID NO. 306); UGCGGCCUCUGUUUGAUUU (SEQ. ID NO. 307); GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);
- 30 GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).
11. The siRNA molecule of claim 10, wherein said siRNA molecule comprises a sequence selected from the group consisting of

GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301);
GAAGUACAUCCAUUAUAAG (SEQ. ID NO. 302);
GUACGACAACCGGGAGAU (SEQ. ID NO. 303);
AGAUAGUGAUGAAGUACAU (SEQ. ID NO. 304);
5 UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305);
GCAUGCGGCCUCUGUUUGA (SEQ. ID NO. 306);
UGCGGCCUCUGUUUGAUUU (SEQ. ID NO. 307);
GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);
GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and
10 GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).

12. The siRNA molecule of claim 11, wherein said siRNA molecule comprises GCAUGCGGCCUCUGUUUGA .
13. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a sense strand and an anti-sense strand.
14. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a hairpin.
15. The siRNA molecule of claim 9, wherein said siRNA molecule comprises between 18 and 30 base pairs.
16. A kit for gene silencing comprising at least one siRNA selected from the group consisting of sequences substantially similar to the group consisting of GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301); GAAGUACAUCCAUUAUAAG (SEQ. ID NO. 302); GUACGACAACCGGGAGAU (SEQ. ID NO. 303); AGAUAGUGAUGAAGUACAU (SEQ. ID NO. 304); UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305); GCAUGCGGCCUCUGUUUGA (SEQ. ID NO. 306); UGCGGCCUCUGUUUGAUUU (SEQ. ID NO. 307); GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);

GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and
GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).

- 5 17. A method of gene silencing comprising using the siRNA molecule of claim 10.
18. A method of gene silencing comprising using the siRNA molecule of claim 11.
- 10 19. A kit, wherein said kit is comprised of at least two siRNA, wherein said at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein said first optimized siRNA and said second optimized siRNA are optimized according to one of the following formulas:

15 Formula I = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

Formula II = $-(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$;

20 Formula III = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3$;

Formula IV = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

25 Formula V = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$;

Formula VI = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$;

30 Formula VII = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2$;

wherein in Formulas I – VII:

- $AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at positions 15 –19;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;
 GC = the number of G and C bases in the entire sense strand;
 $Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;
 $A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;
 $A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;
 $U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+$

$$(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$$

wherein

- 5 $A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
10 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
15 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

- $C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
20 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
25 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

- $G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
30 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;

5 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;

$U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;

$U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;

$U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;

$U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;

10 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

$U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;

$U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;

$U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;

$U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

15

GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand
or within positions 15 –18 if the sense strand is only 18 base pairs in length;

GC_{total} = the number of G and C bases in the sense strand;

$T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs,

20 otherwise its value is 0; and

X = the number of times that the same nucleotide repeats four or more times in a row.

Abstract

Efficient sequence specific gene silencing is possible through the use of siRNA technology. By selecting particular siRNAs by rationale design, one can maximize the generation of an effective gene silencing reagent, as well as methods for
5 silencing genes.

10

Figure 1

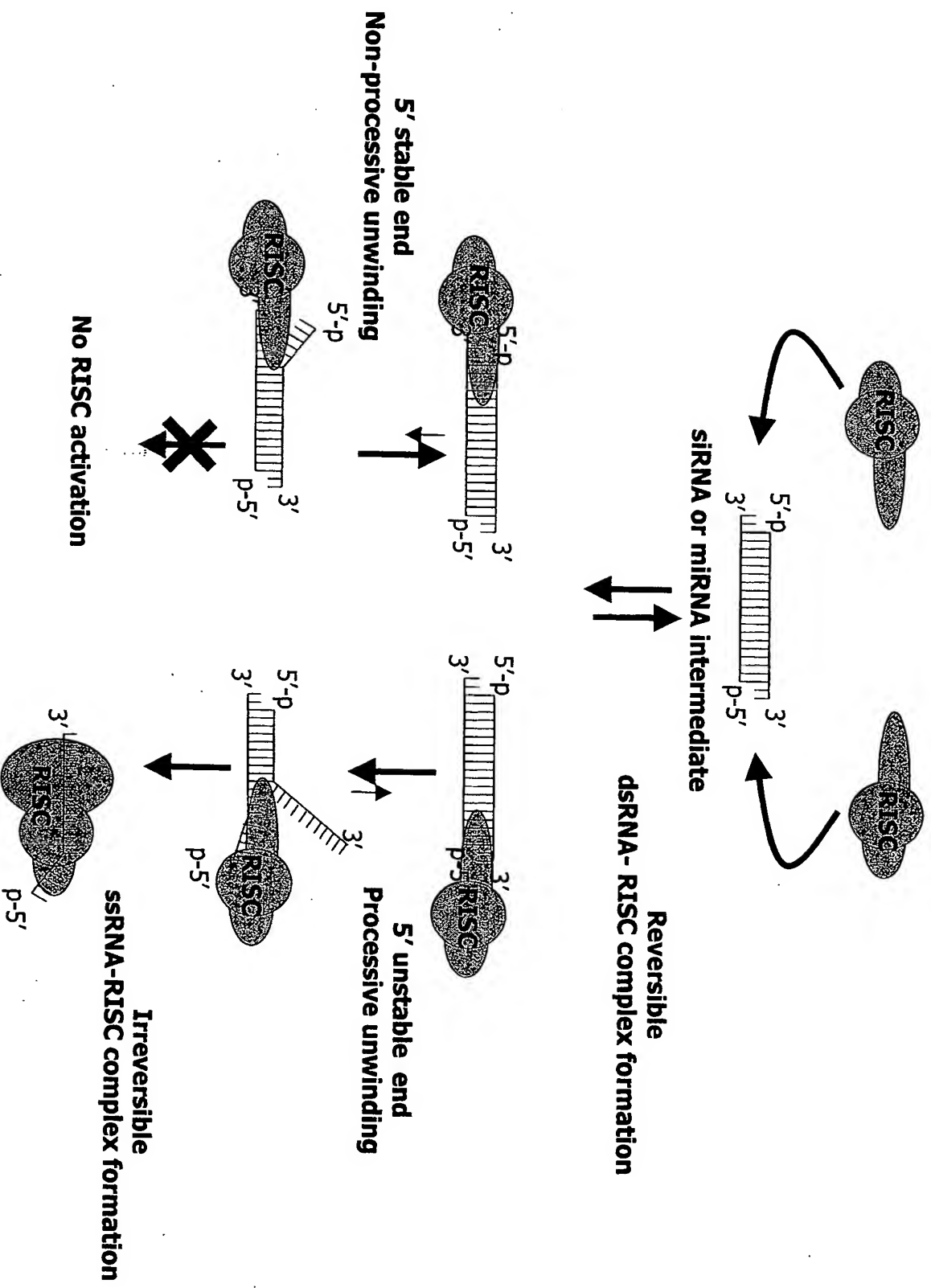


Figure 2

siRNA panel (270)

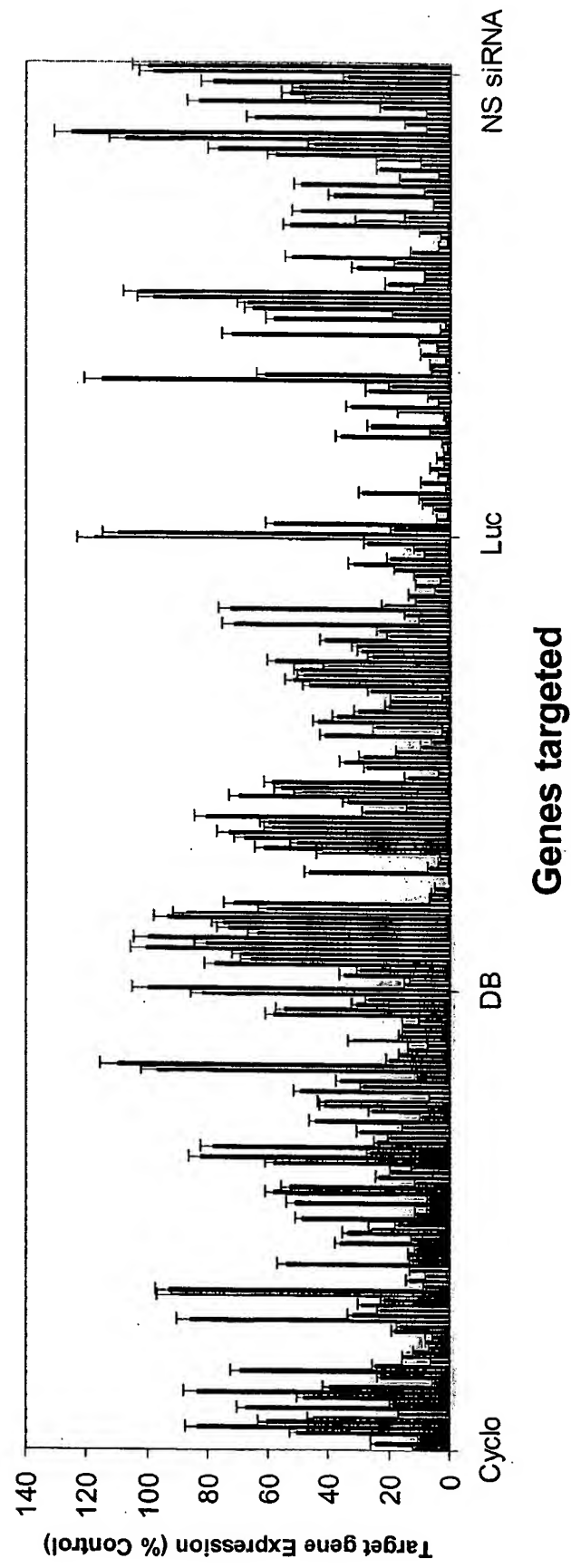
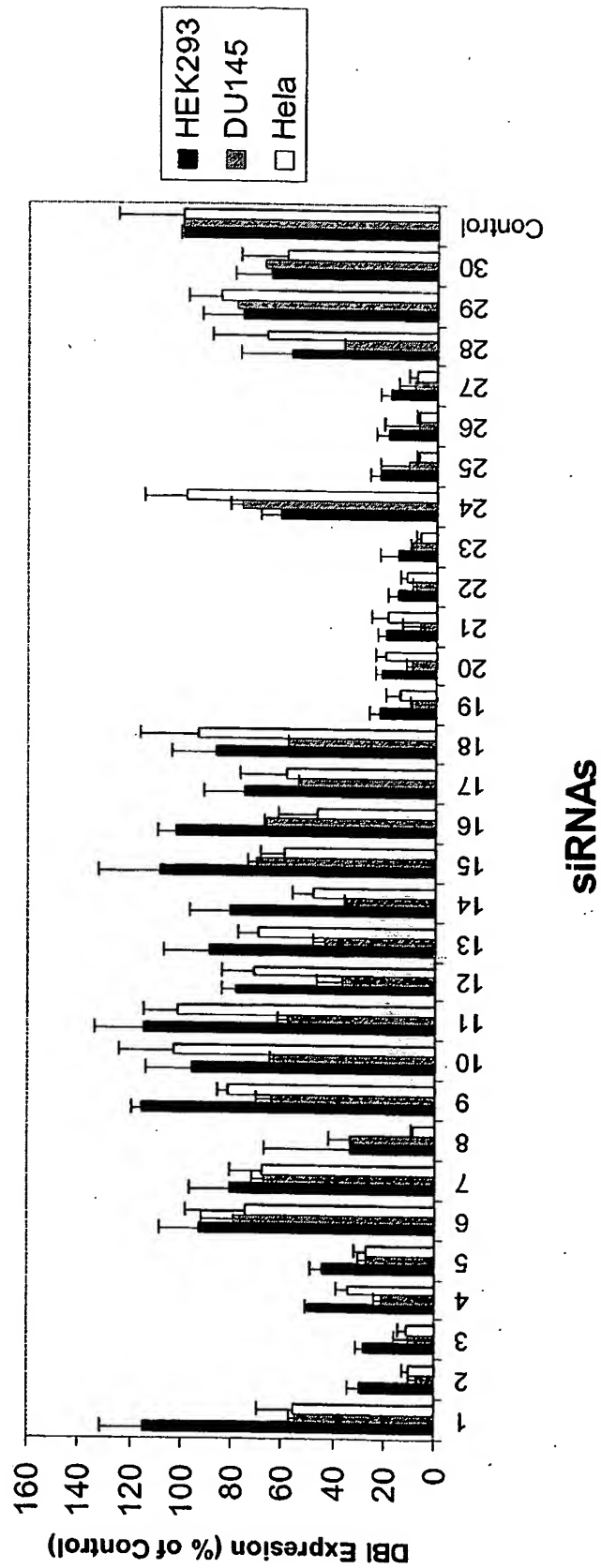


Figure 3a

siRNA functionality is independent from the cell line



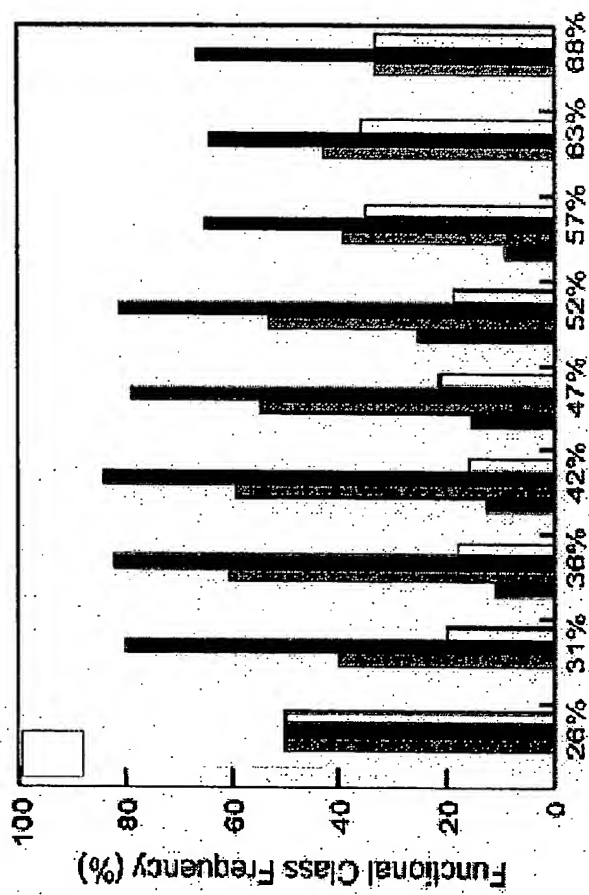


Figure 3b

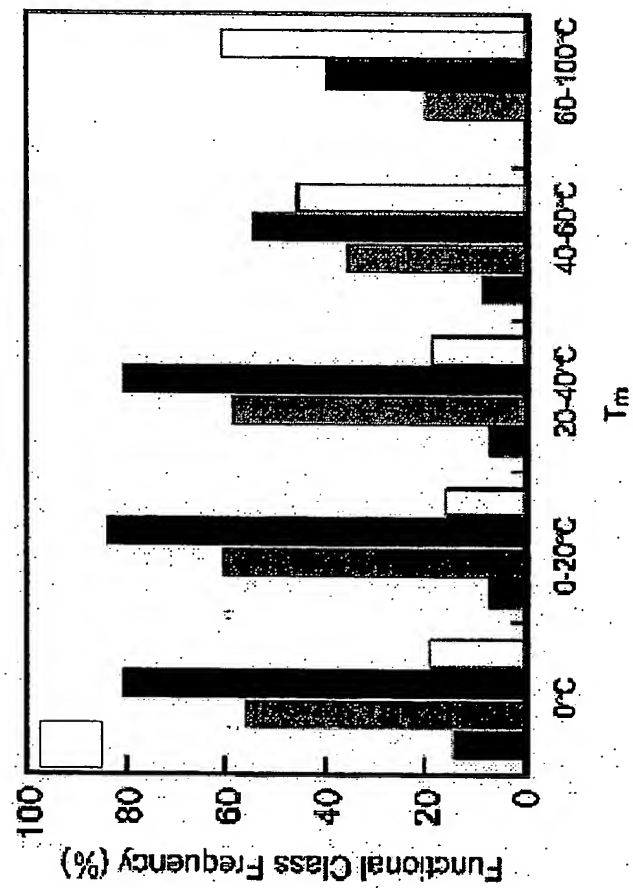


Figure 3c

Figure 4

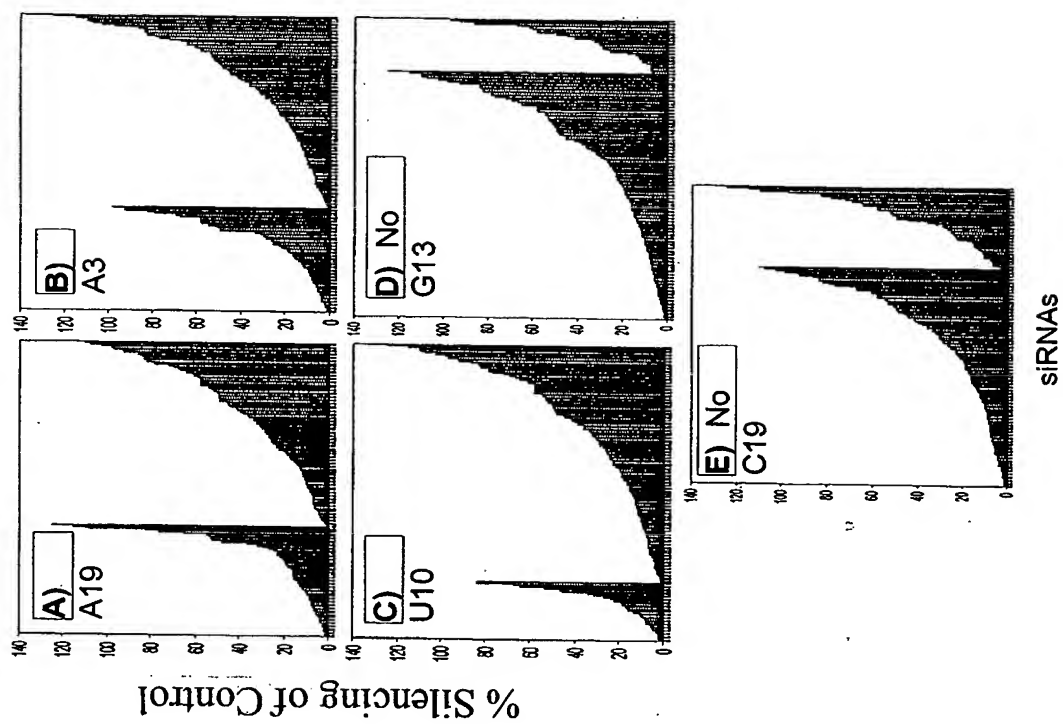


Figure 5A

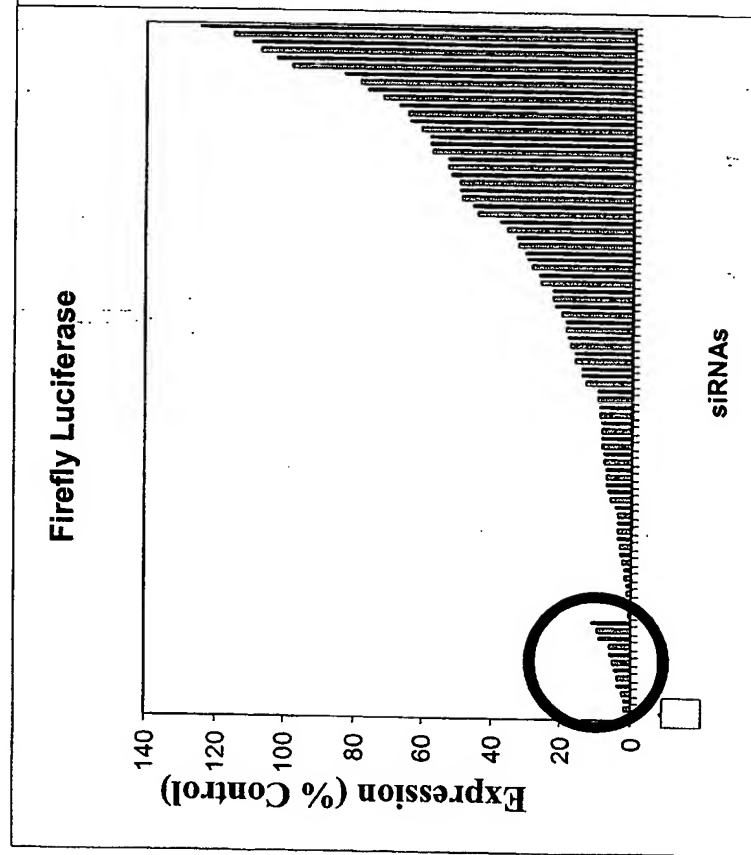


Figure 5B

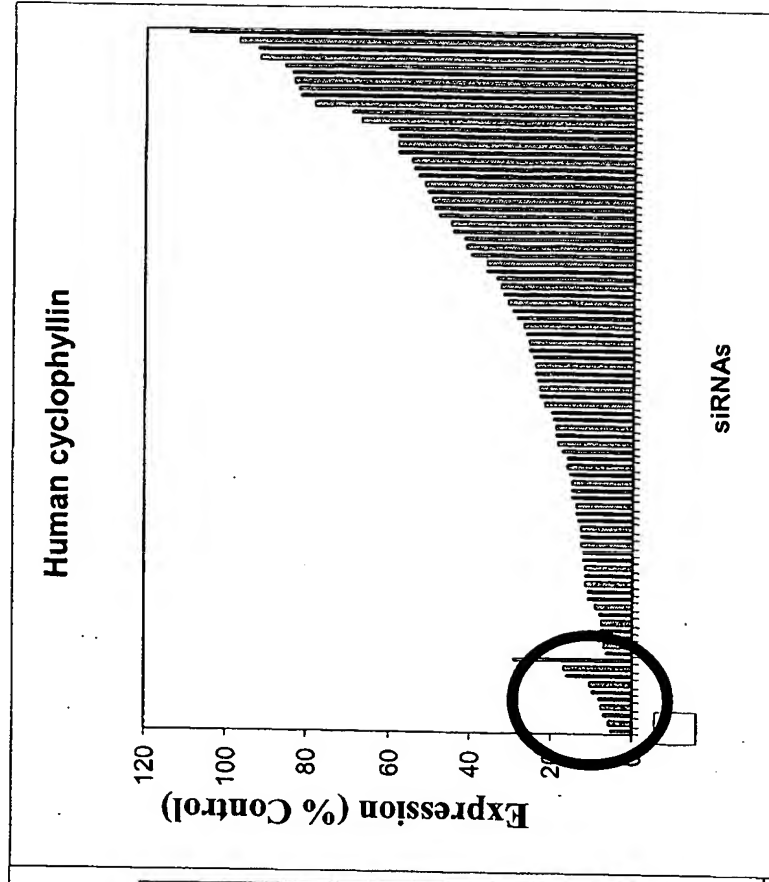


Figure 6a

Differential internal stability

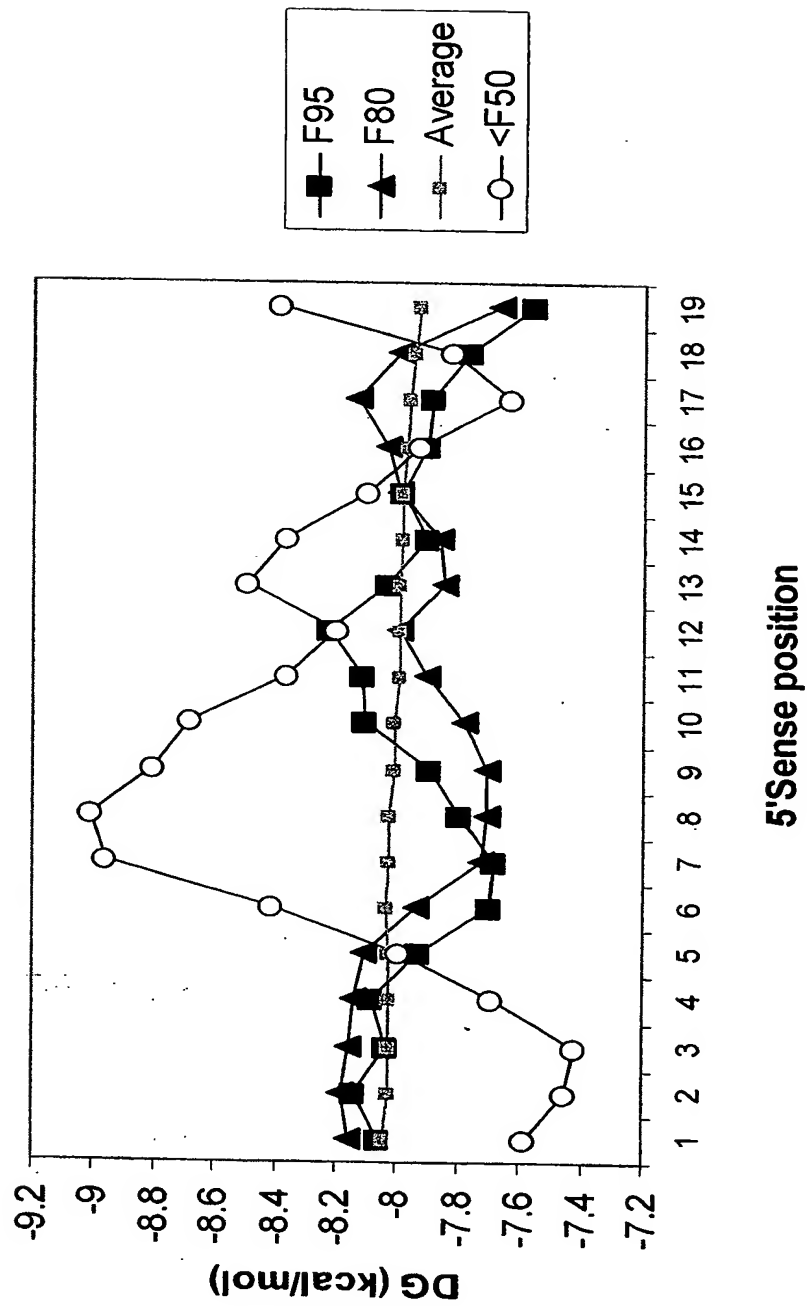


Figure 6b

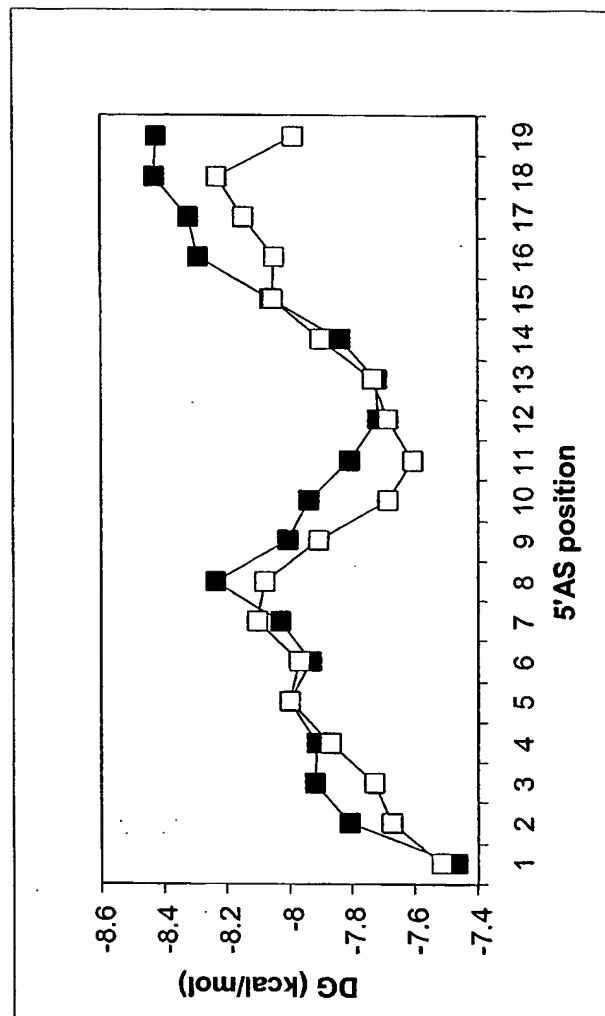
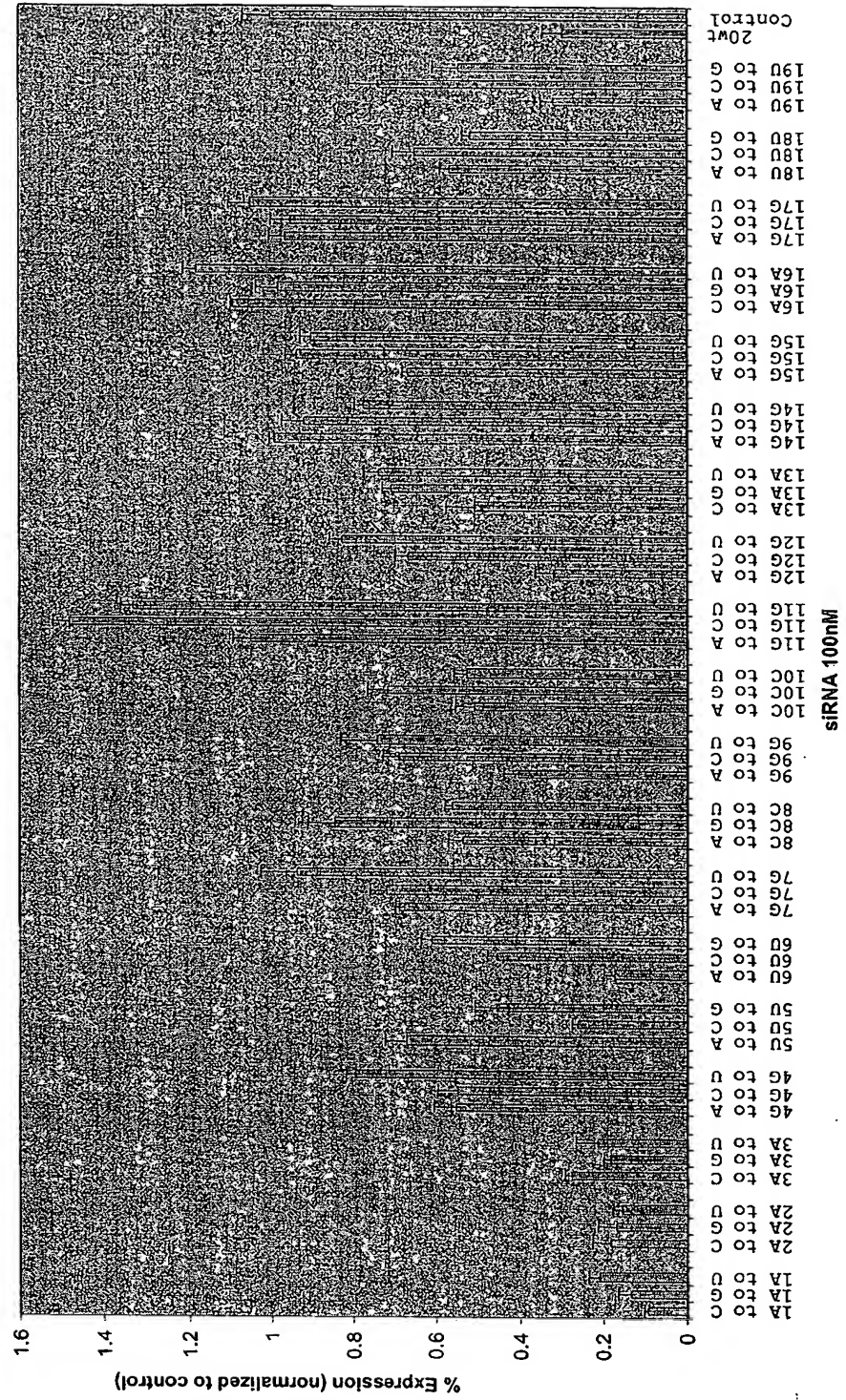
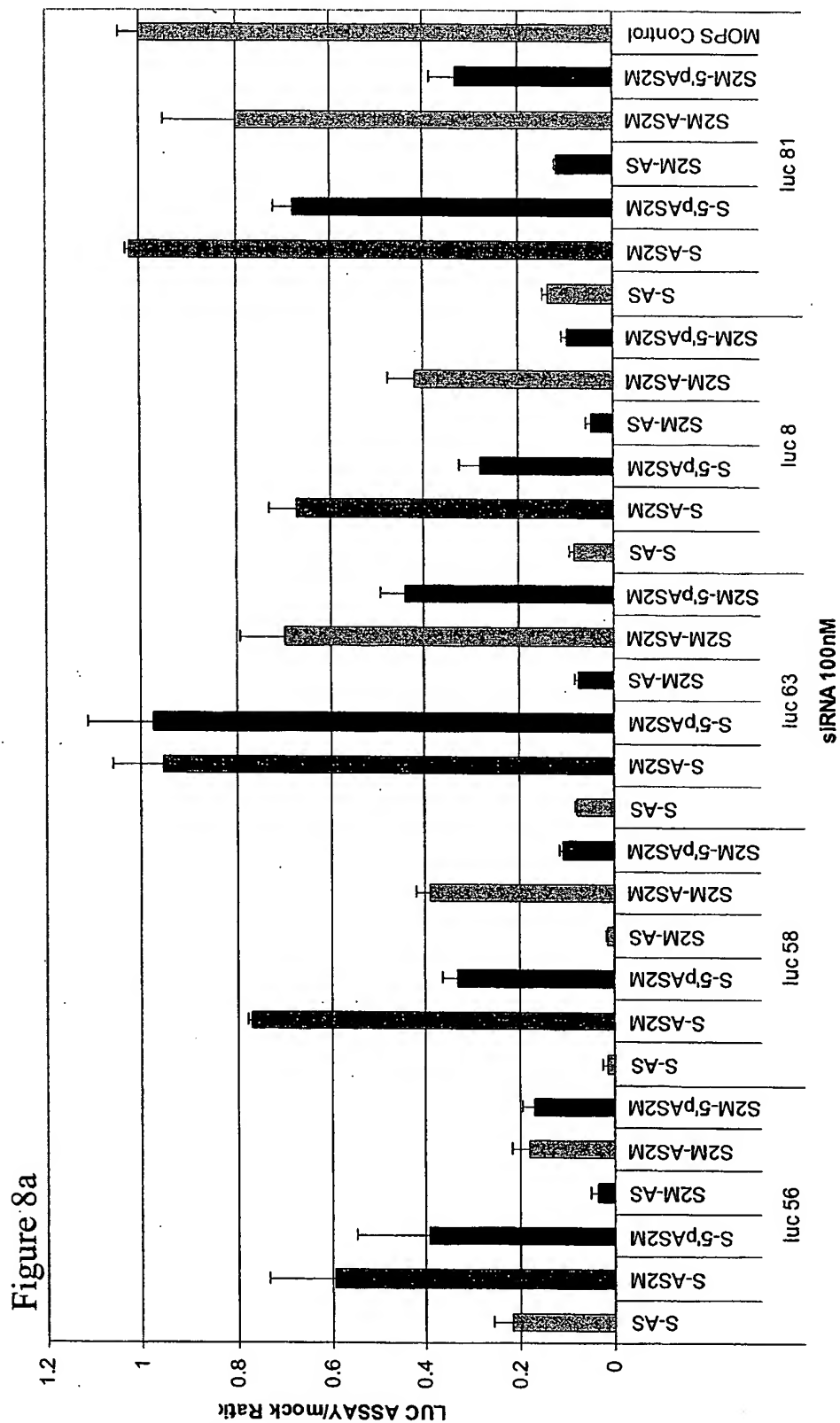


Figure 7



TARGET Screen Normalized LUC ASSAY 293 cells



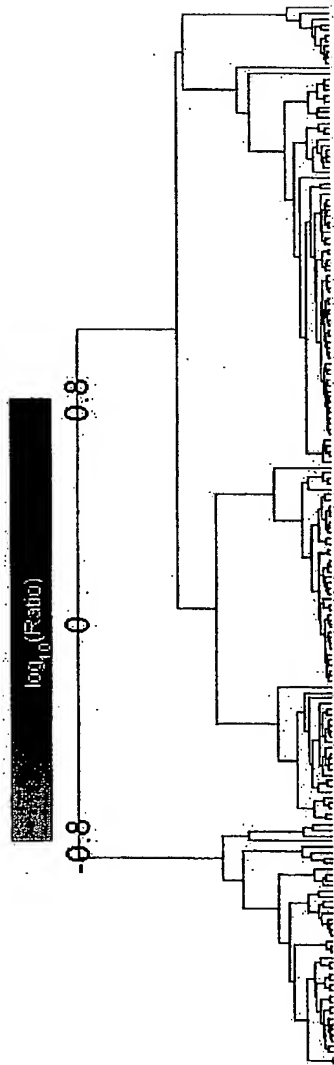
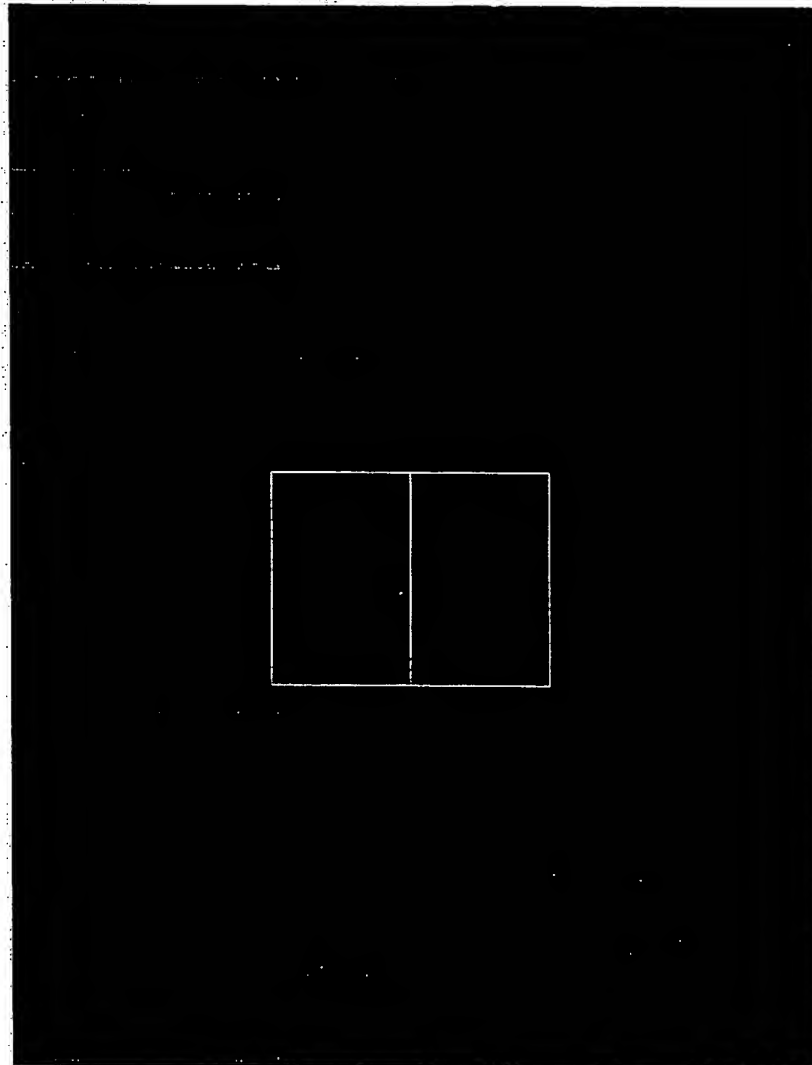


Figure 8b

IGF1R-3

Sense strand modified

IGF1R-3



CLUSTER ANALYSIS SUMMARY:
 12-Aug-2006 16:37
 CG from Affet 1637
 Clustering parameters: algorithm = p-value, similarity measure = correlation-based, metric type = link
 Statistics: 133 genes in 6 experiments selected from 23661 genes in 6 experiments

Figure 9

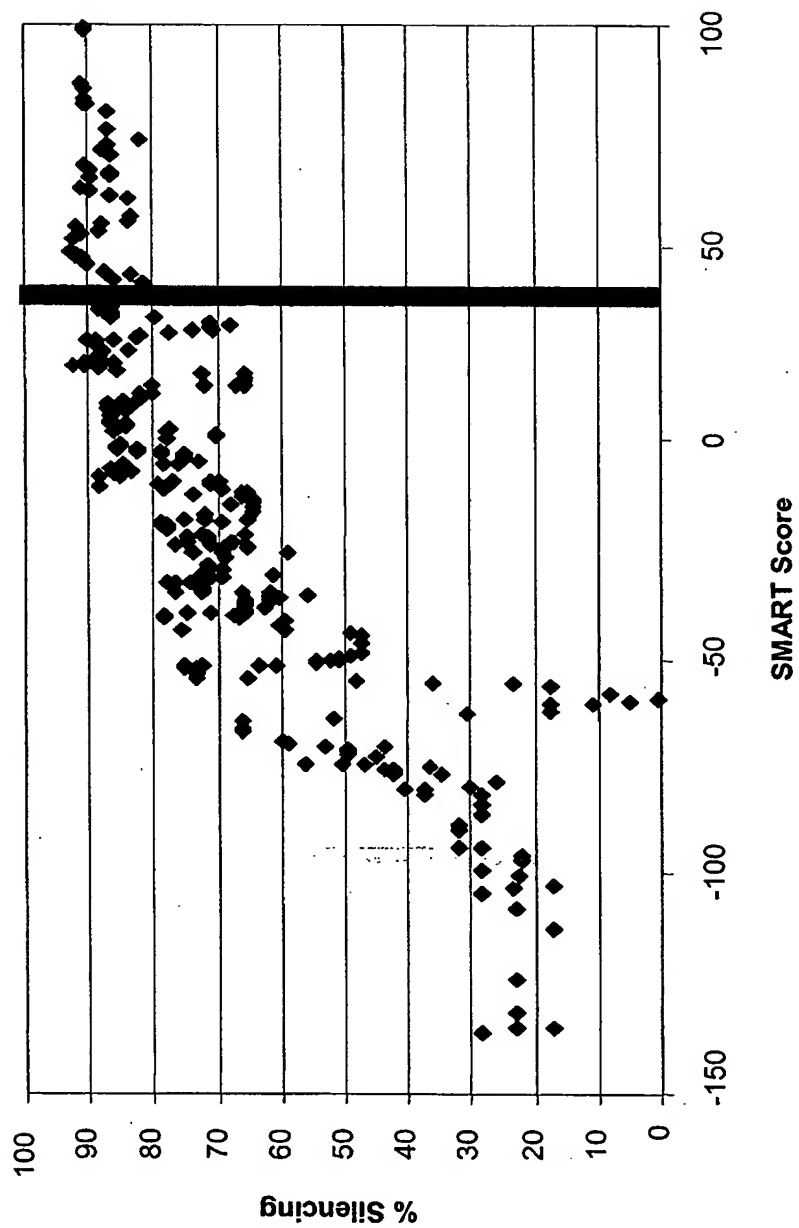


Figure 10a-f

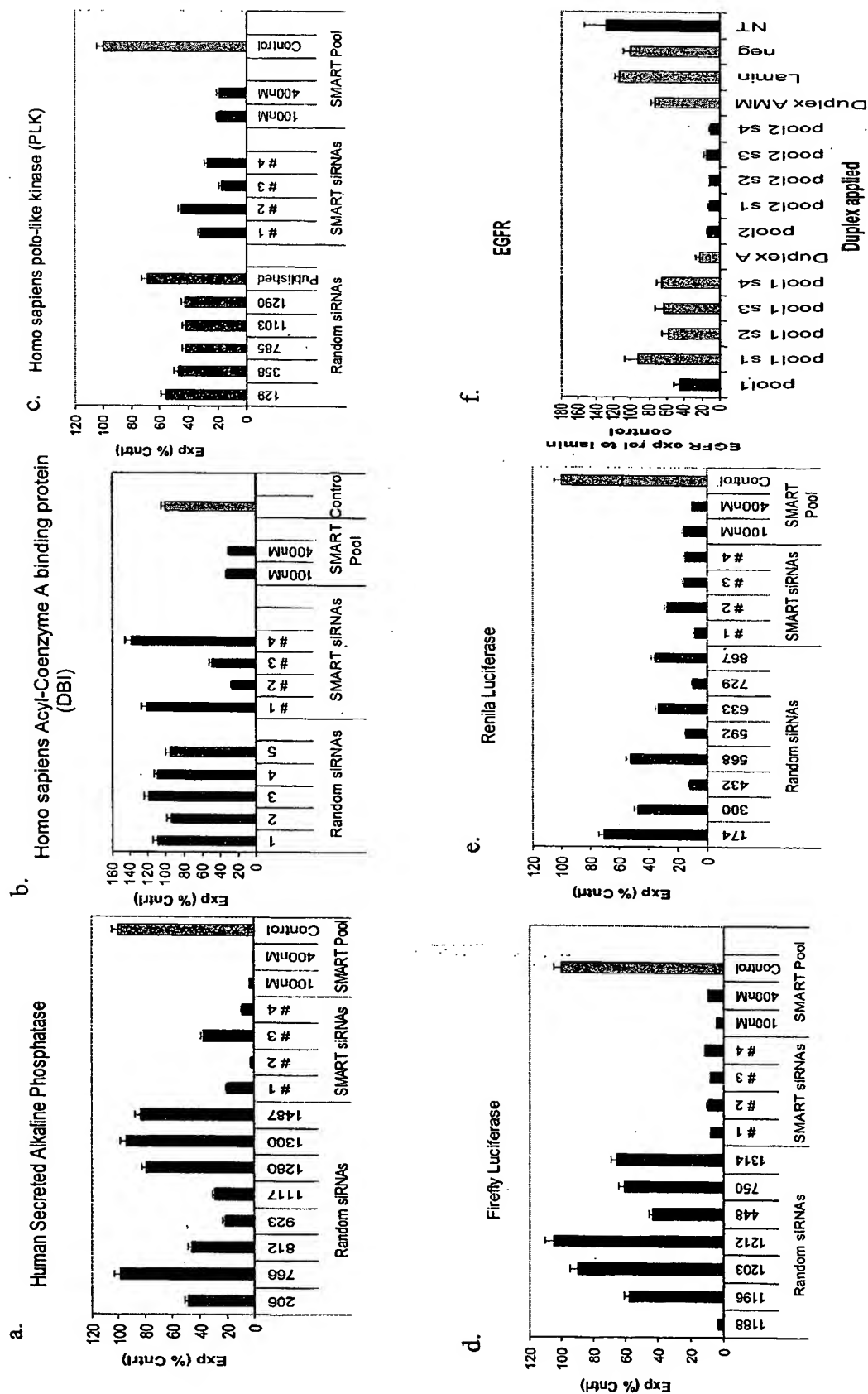


Figure 11

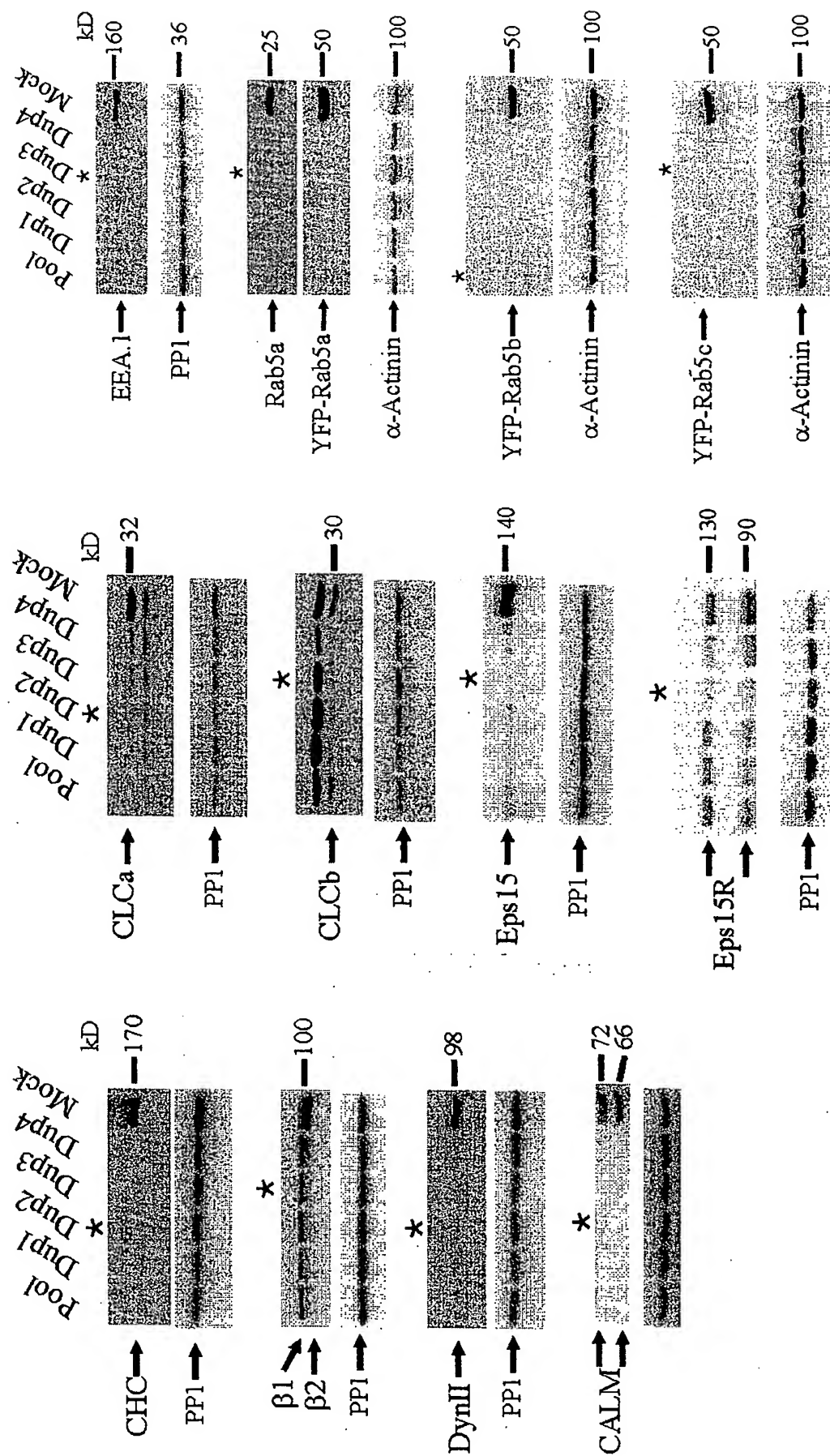


Figure 12

Rational selection validation

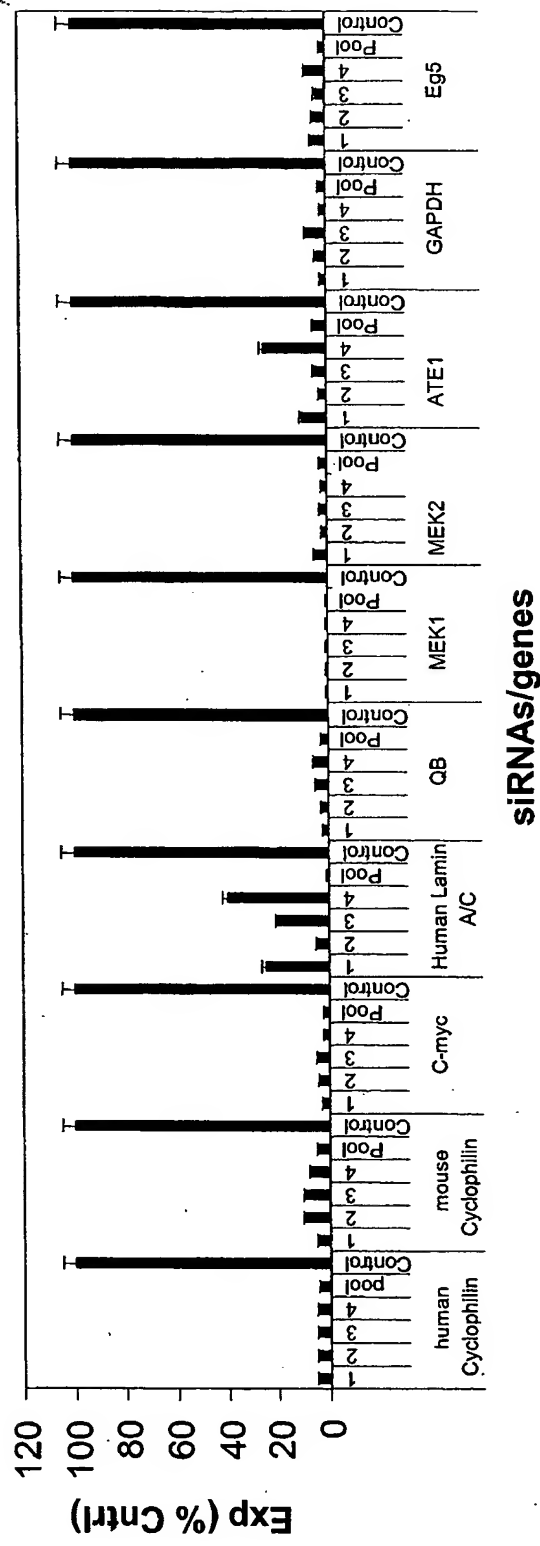
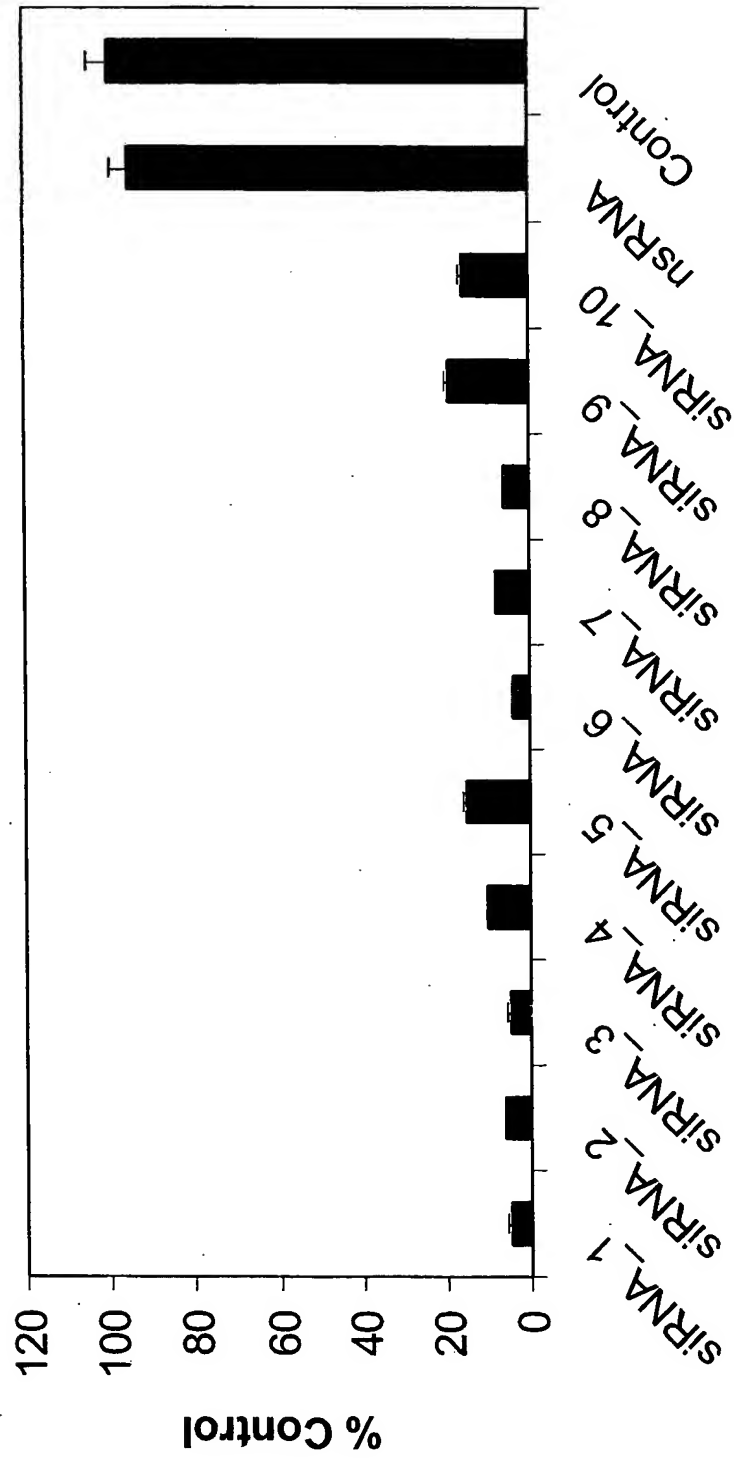


Figure 13 Sequences of top Bcl2

siRNA 1	GGGAGAUAGUGAAGUA
siRNA 2	GAAGUACAUCCAUUAAG
siRNA 3	GUACGACAACCGGAGUA
siRNA 4	AGAUAGUGAUGAAGUACAU
siRNA 5	UGAAGACUCUGCUCAGUUU
siRNA 6	GCAUGCGGCCUCUGUUUGA
siRNA 7	UGC GGCCUCUGUUUGAUUU
siRNA 8	GAGAUAGUGAAGUAACA
siRNA 9	GGAGAUAGUGAAGUAC
siRNA 10	GAAGACUCUGCUCAGUUUG

Figure 14

**Bcl-2 knockdown by 10 rationally designed siRNAs at
100 nM concentration**



Reporter gene individual siRNAs walk

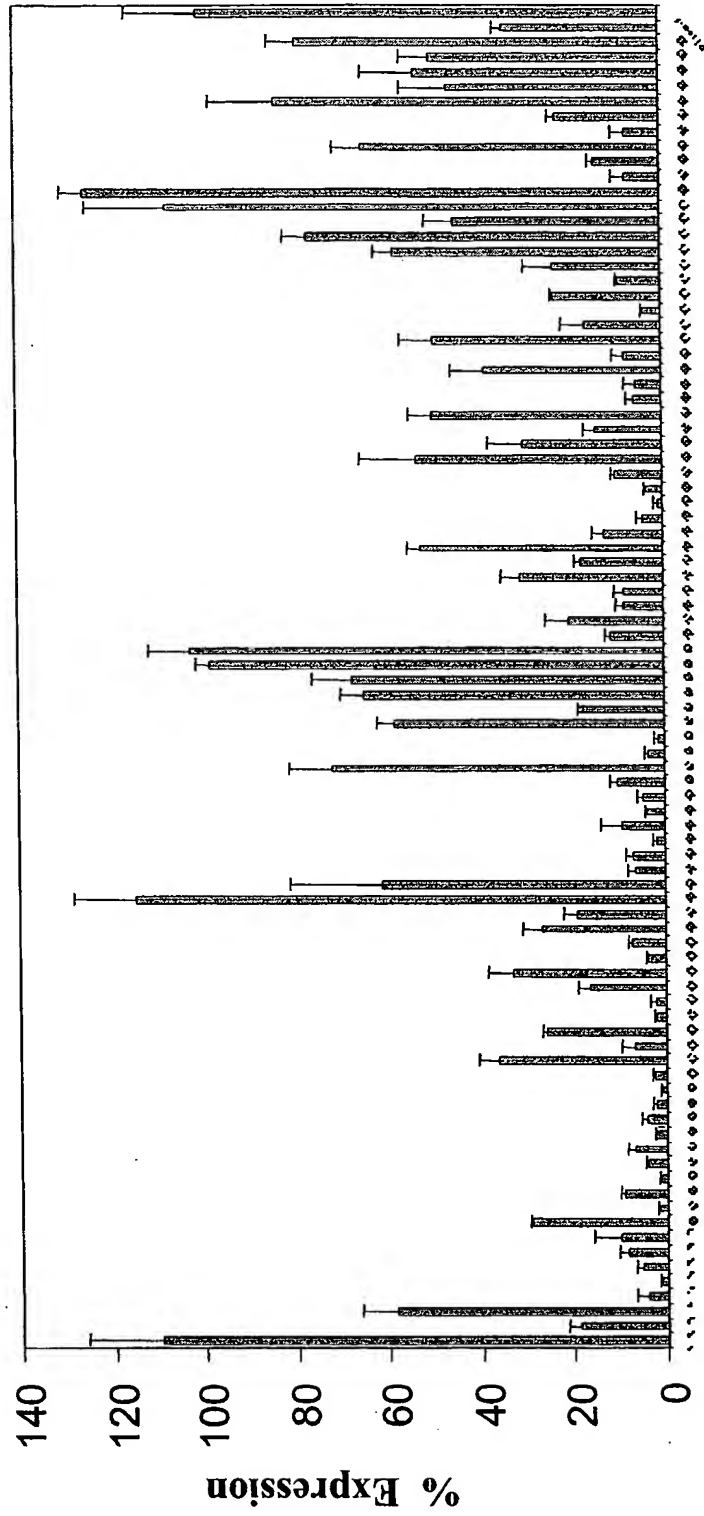


Figure 15

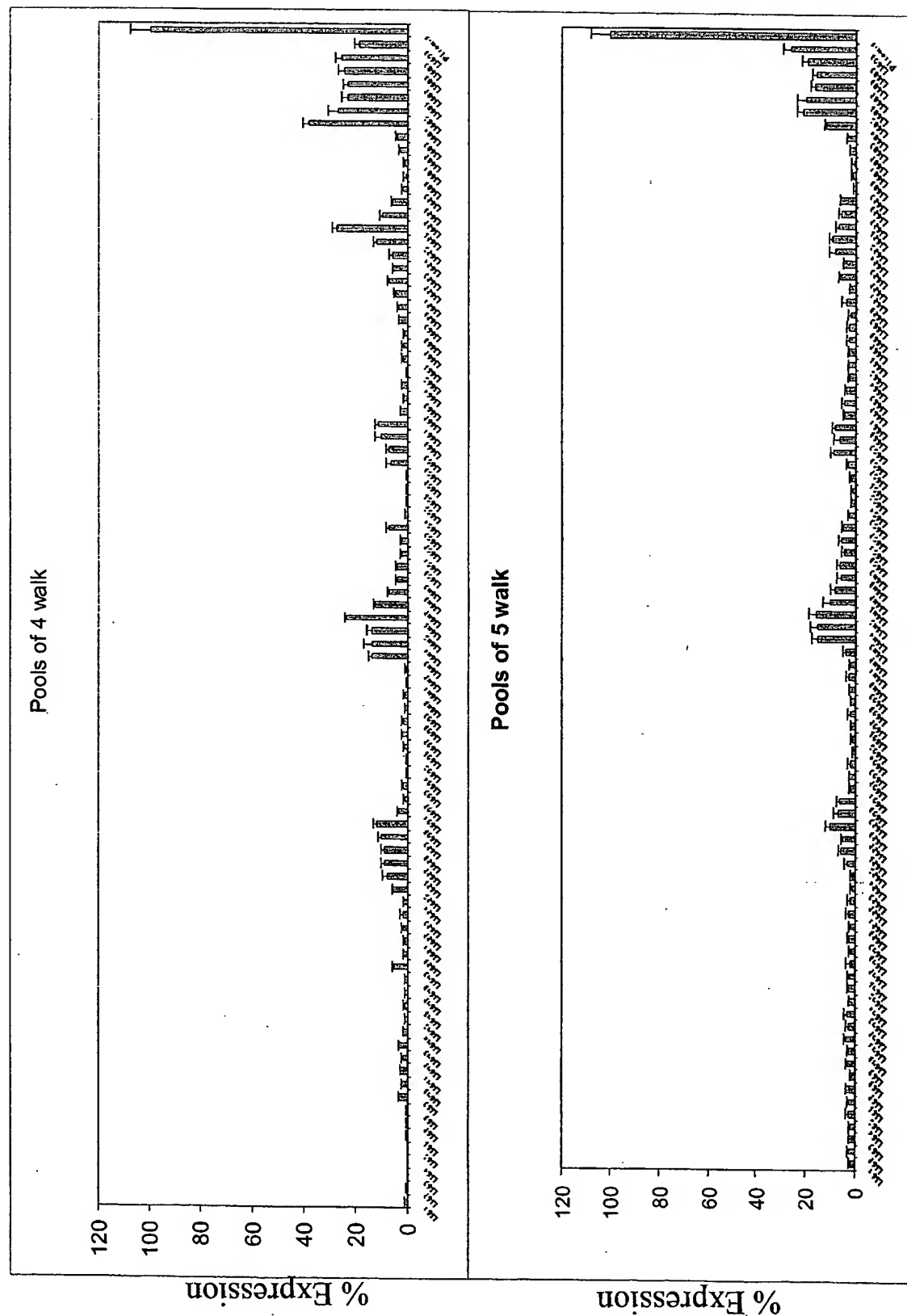


Figure 17

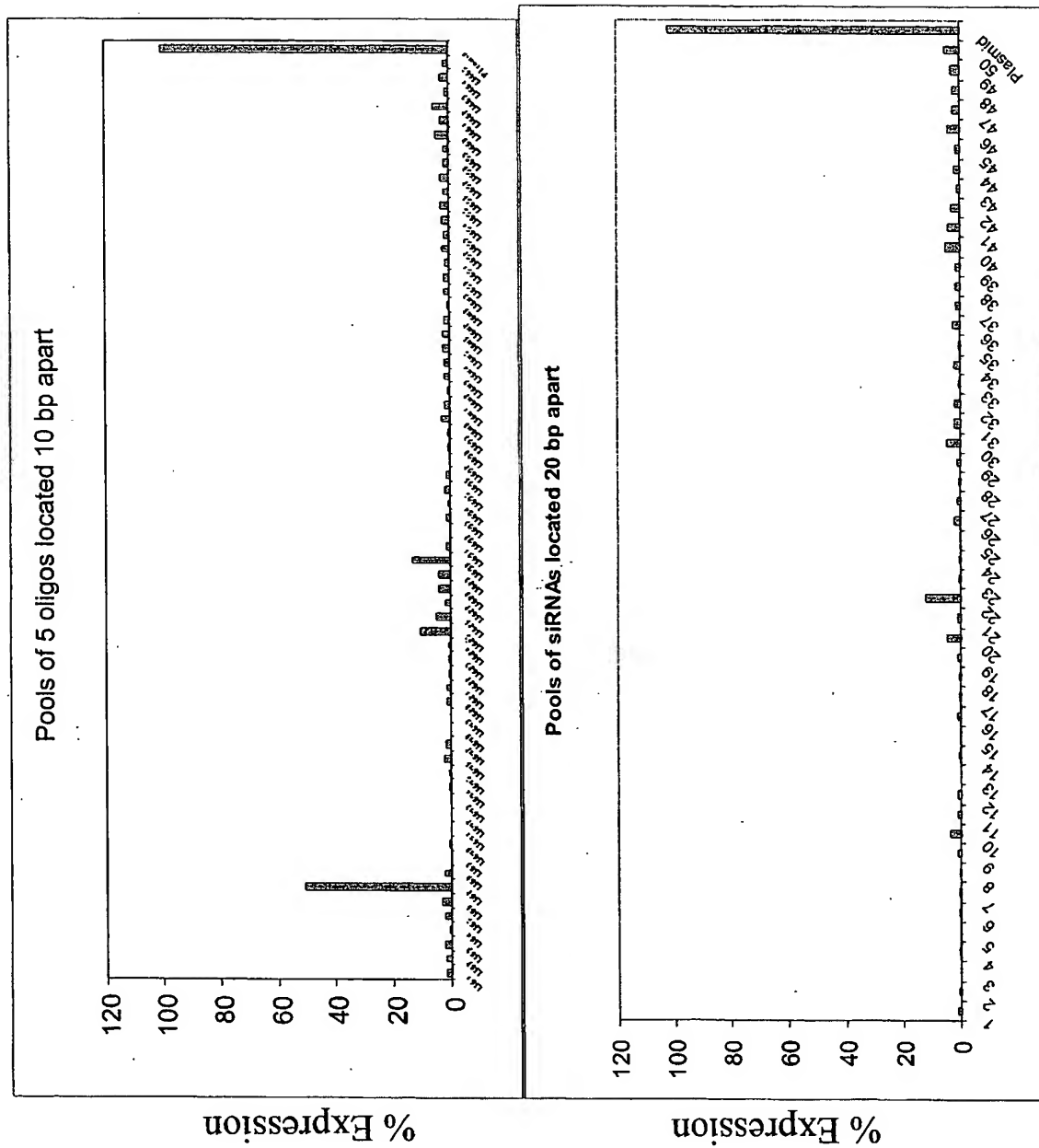
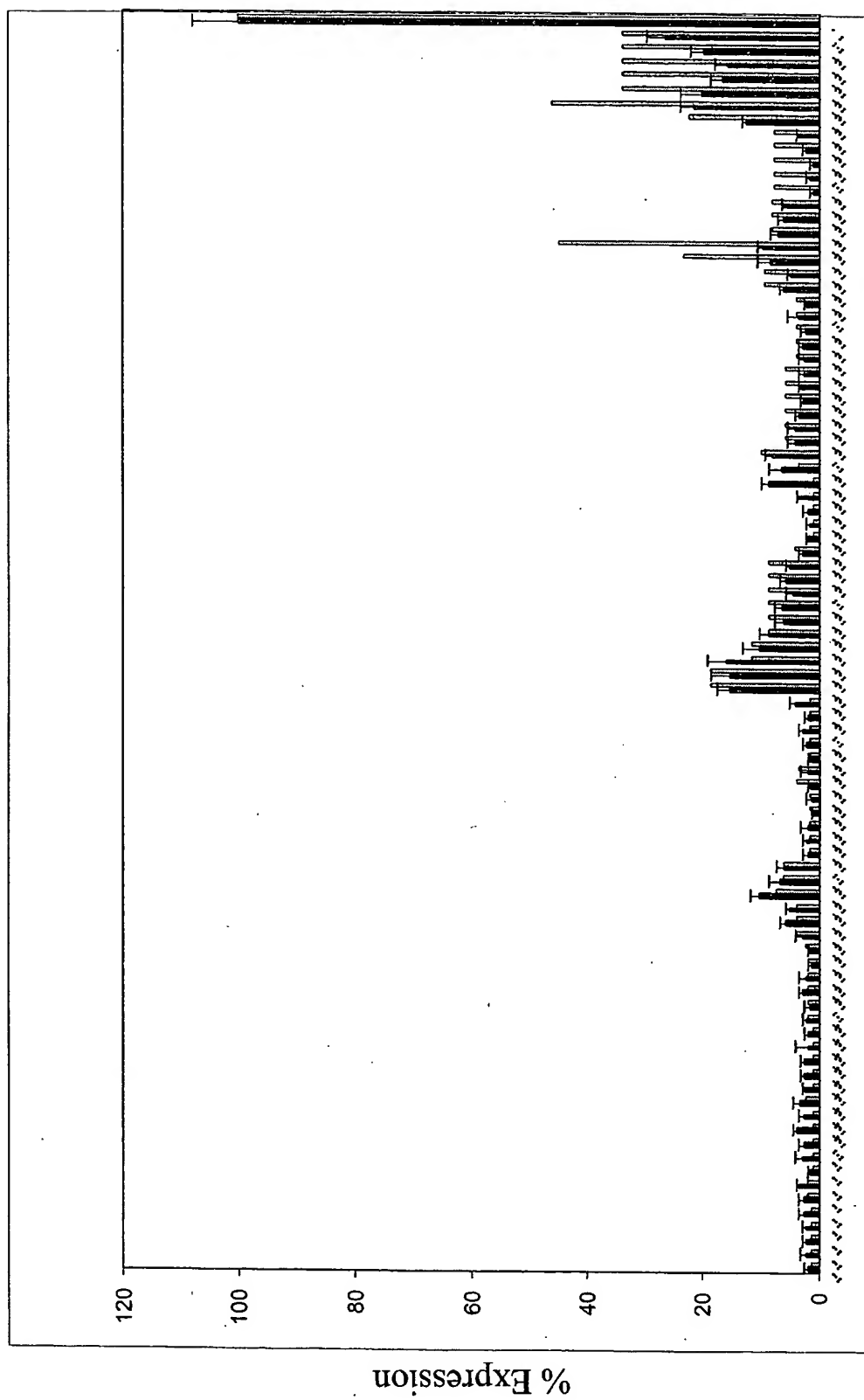


Figure 18



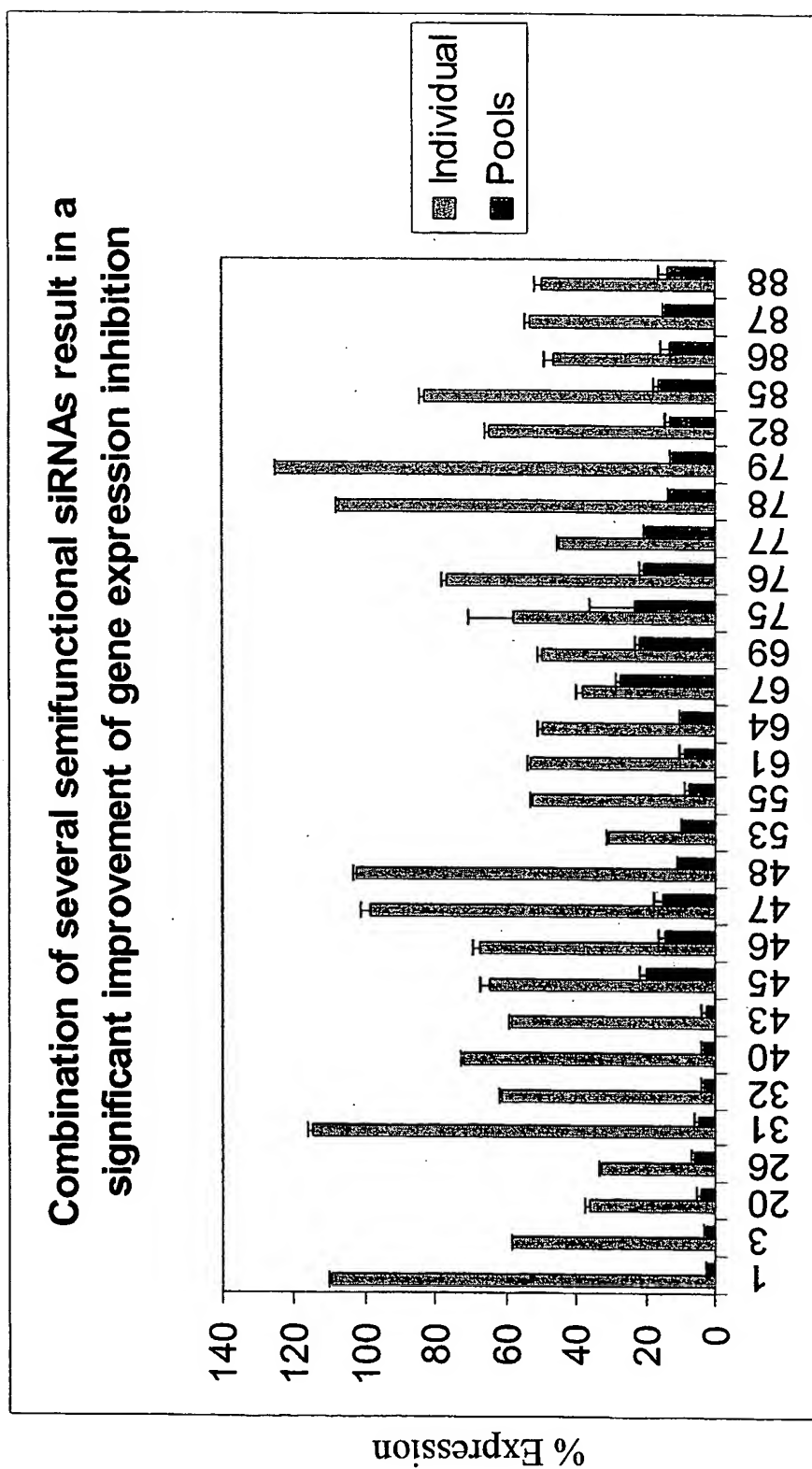


Figure 20

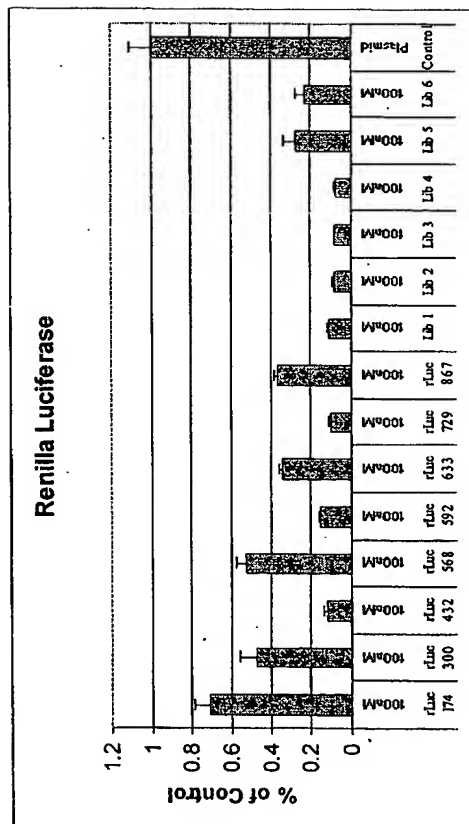
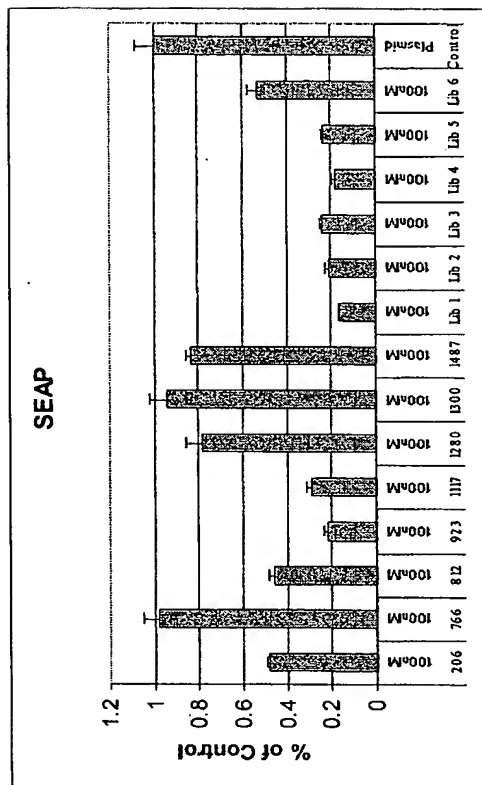
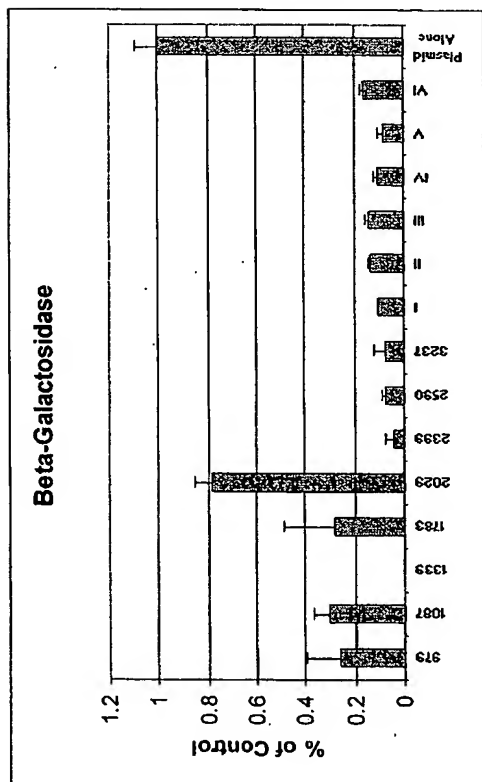


Figure 21

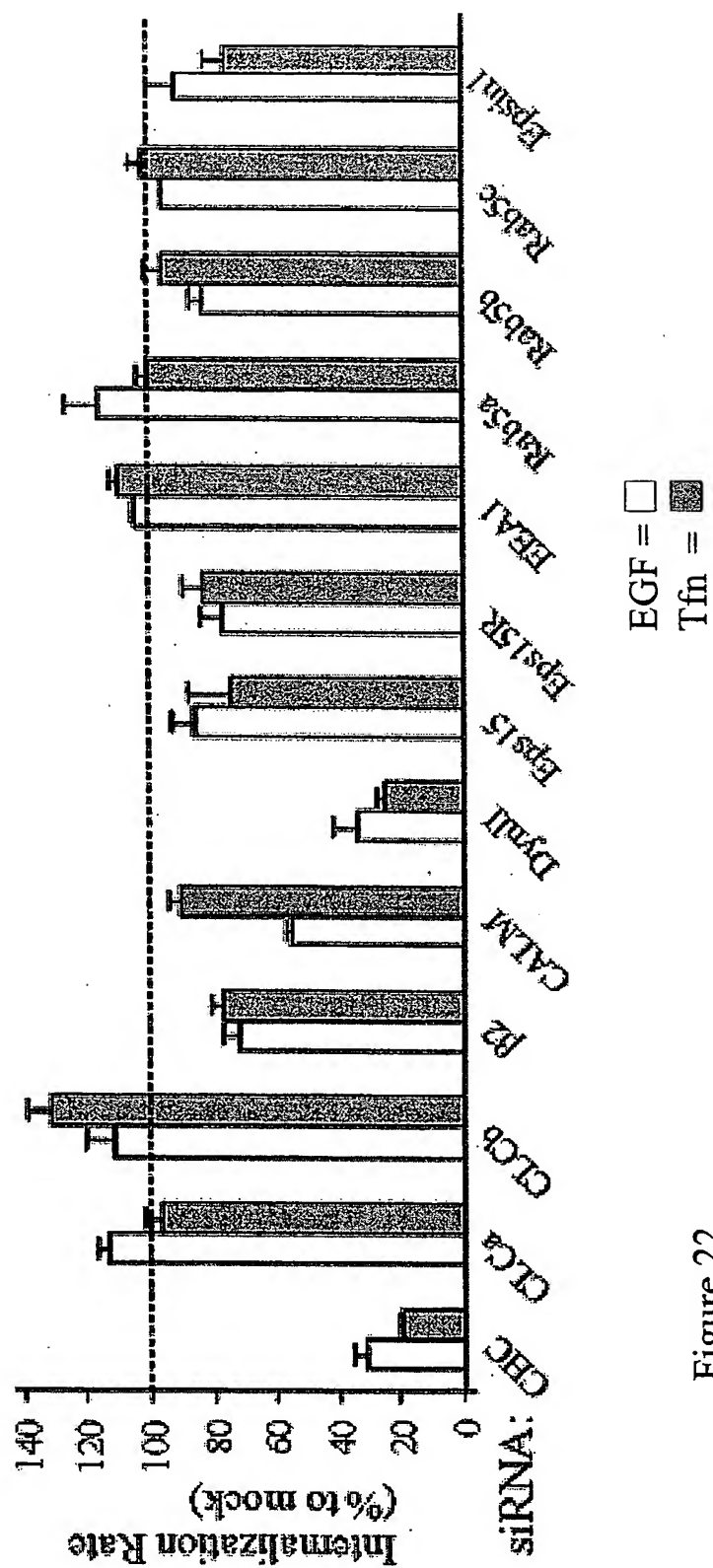


Figure 22

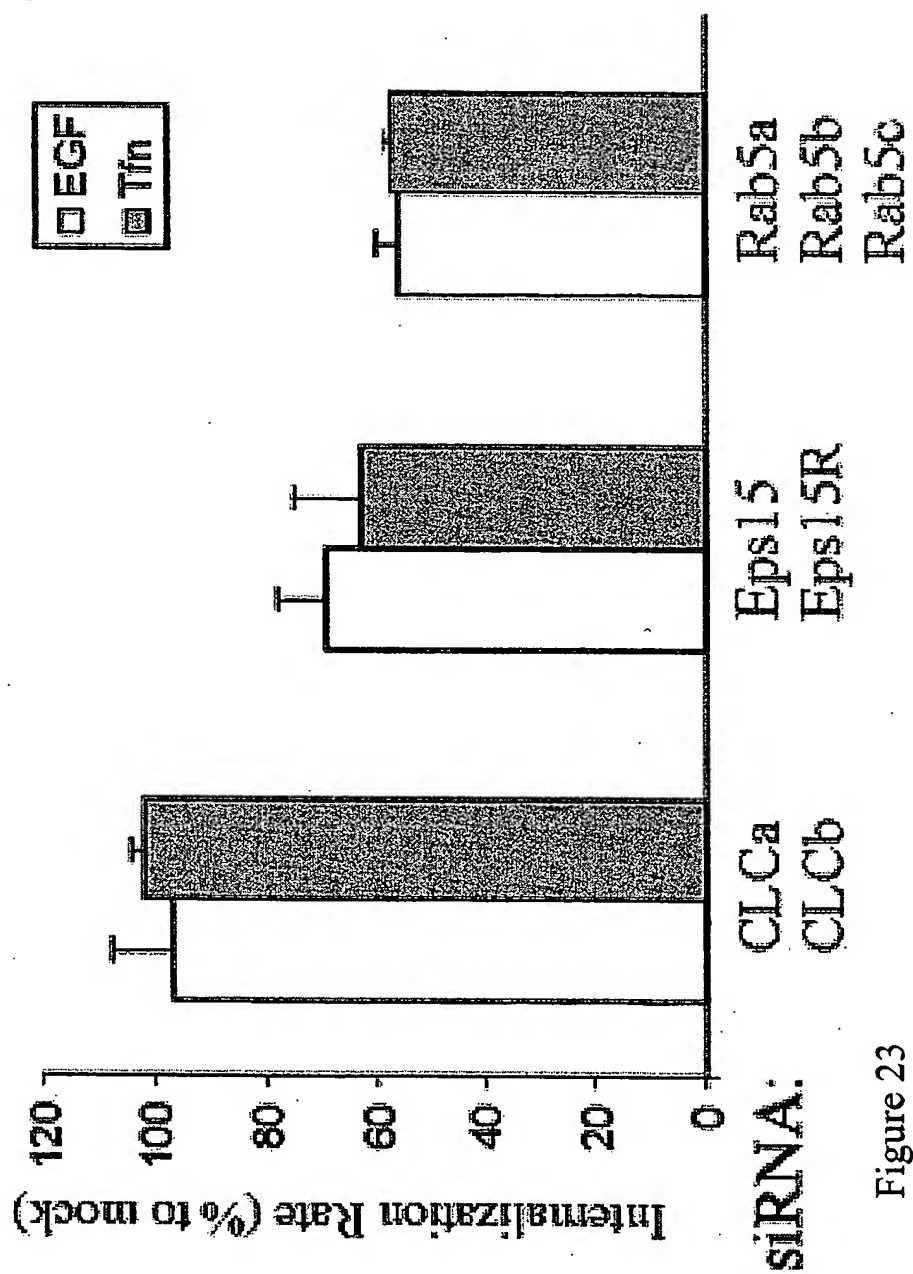


Figure 23

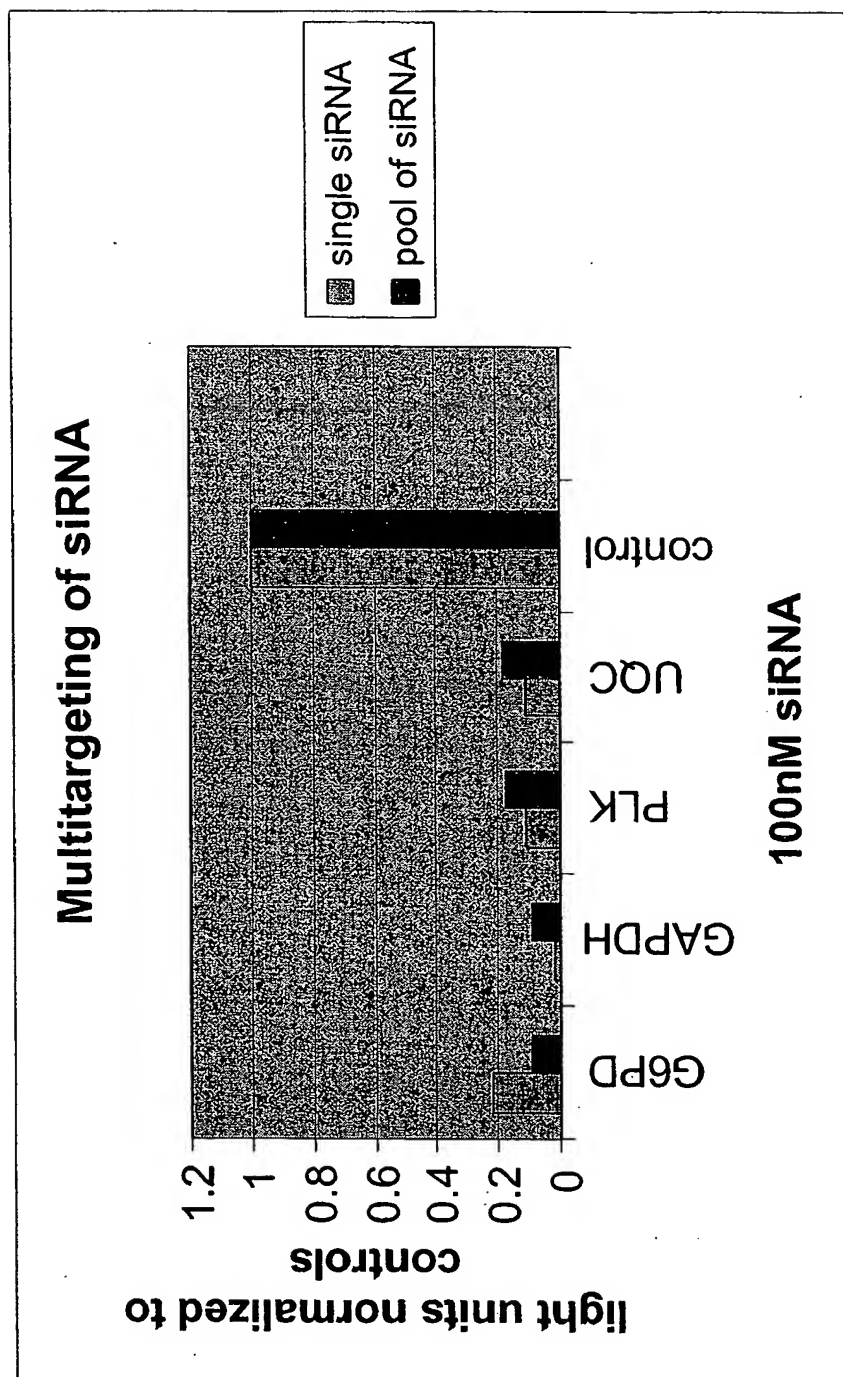
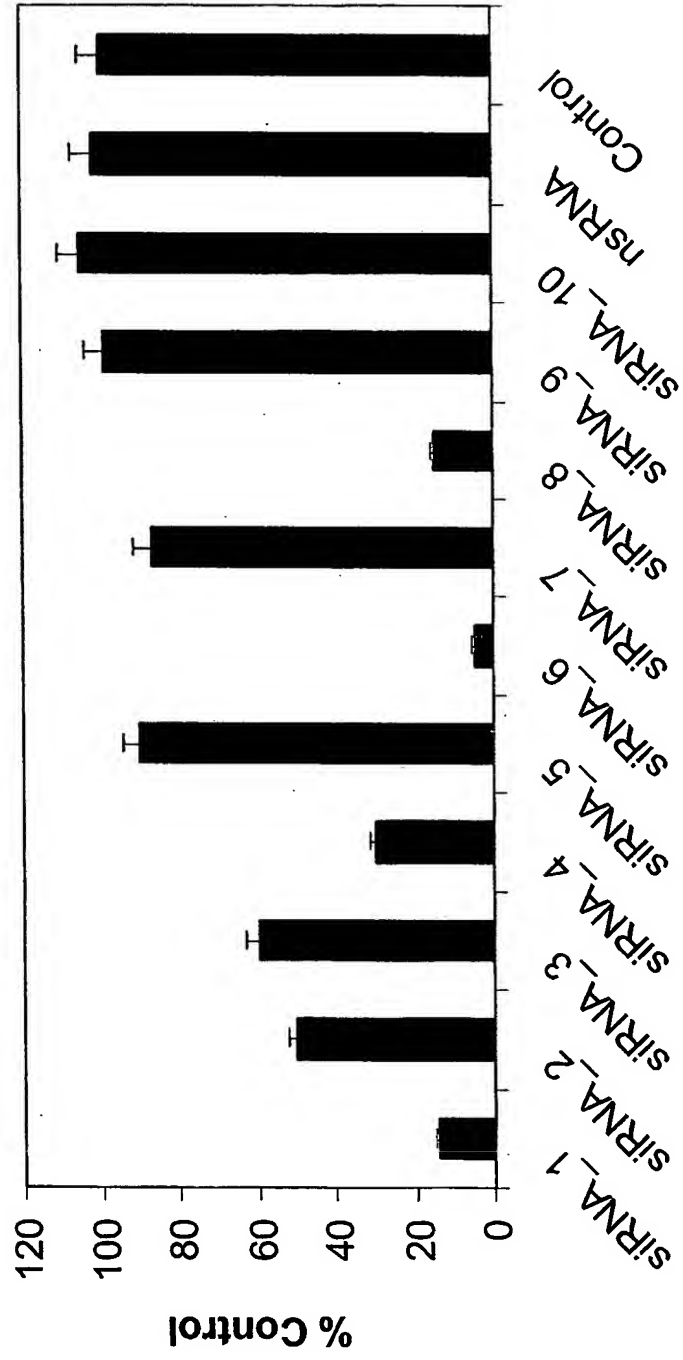


Figure 24

Figure 25

**Bcl-2 knockdown by 10 rationally designed siRNAs at
300 pM concentration**



ITEM 2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Khvorova *et al.* Examiner: To be assigned
Serial No.: 10/714,333 Group Art Unit: To be assigned
Filed: November 14, 2003
For: Functional and Hyperfunctional siRNA
Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

May 20, 2004

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

PRELIMINARY AMENDMENT

Dear Sir:

Prior to examination on the merits, please enter the following amendments in the above-identified application.

Certificate of Mailing Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited on the date shown below with the United States Postal Service as first class mail with sufficient postage in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

Viola Hyland
(Signature)

Viola Hyland
(Printed Name of Person Signing
Certificate)

May 20, 2004
(Date)

Amendments to the Specification:

Please replace Table III, spanning pages 57-63, with the following amended Table III:

[[**TABLE III**]]

<u>TABLE III</u>			
Cyclo	1	SEQ. ID 0032	GUUCCAAAAACAGUGGAUA
Cyclo	2	SEQ. ID 0033	UCCAAAAACAGUGGAUAAU
Cyclo	3	SEQ. ID 0034	CAAAAACAGUGGAUAAUUU
Cyclo	4	SEQ. ID 0035	AAAACAGUGGAUAAUUUUG
Cyclo	5	SEQ. ID 0036	AACAGUGGAUAAUUUUGUG
Cyclo	6	SEQ. ID 0037	CAGUGGAUAAUUUUGUGGC
Cyclo	7	SEQ. ID 0038	GUGGAUAAUUUUGUGGCCU
Cyclo	8	SEQ. ID 0039	GGAUAAUUUUGUGGCCUUA
Cyclo	9	SEQ. ID 0040	AUAAUUUUGUGGCCUAGC
Cyclo	10	SEQ. ID 0041	AAUUUUGUGGCCUAGCUA
Cyclo	11	SEQ. ID 0042	UUUUGUGGCCUAGCUACA
Cyclo	12	SEQ. ID 0043	UUGUGGCCUAGCUACAGG
Cyclo	13	SEQ. ID 0044	GUGGCCUAGCUACAGGAG
Cyclo	14	SEQ. ID 0045	GGCCUAGCUACAGGAGAG
Cyclo	15	SEQ. ID 0046	CCUAGCUACAGGAGAGAA
Cyclo	16	SEQ. ID 0047	UUAGCUACAGGAGAGAAAG
Cyclo	17	SEQ. ID 0048	AGCUACAGGAGAGAAAGGA
Cyclo	18	SEQ. ID 0049	CUACAGGAGAGAAAGGAUU
Cyclo	19	SEQ. ID 0050	ACAGGAGAGAAAGGAUUUG
Cyclo	20	SEQ. ID 0051	AGGAGAGAAAGGAUUUGGC
Cyclo	21	SEQ. ID 0052	GAGAGAAAGGAUUUGGCUA
Cyclo	22	SEQ. ID 0053	GAGAAAGGAUUUGGCUACA
Cyclo	23	SEQ. ID 0054	GAAAGGAUUUGGCUACAAA
Cyclo	24	SEQ. ID 0055	AAGGAUUUGGCUACAAAAA
Cyclo	25	SEQ. ID 0056	GGAUUUGGCUACAAAAACA
Cyclo	26	SEQ. ID 0057	AUUUGGCUACAAAAACAGC
Cyclo	27	SEQ. ID 0058	UUGGCUACAAAAACAGCAA

Applicants: Khvorova *et al.*
 Serial No.: 10/714,333
 Filing Date: November 14, 2003
 Preliminary Amendment
 May 20, 2004
 Page 3 of 42

Cyclo	28	SEQ. ID 0059	GGCUACAAAAACAGCAAU
Cyclo	29	SEQ. ID 0060	CUACAAAAACAGCAAUUC
Cyclo	30	SEQ. ID 0061	ACAAAAACAGCAAUCCA
Cyclo	31	SEQ. ID 0062	AAAAACAGCAAUCCAUC
Cyclo	32	SEQ. ID 0063	AAACAGCAAUCCAUCGU
Cyclo	33	SEQ. ID 0064	ACAGCAAUCCAUCGUGU
Cyclo	34	SEQ. ID 0065	AGCAAUCCAUCGUGUAA
Cyclo	35	SEQ. ID 0066	CAAUCCAUCGUGUAAUC
Cyclo	36	SEQ. ID 0067	AAUCCAUCGUGUAAUCAA
Cyclo	37	SEQ. ID 0068	UCCAUCGUGUAAUCAAGG
Cyclo	38	SEQ. ID 0069	CCAUCGUGUAAUCAAGGAC
Cyclo	39	SEQ. ID 0070	AUCGUGUAAUCAAGGACUU
Cyclo	40	SEQ. ID 0071	CGUGUAAUCAAGGACUUA
Cyclo	41	SEQ. ID 0072	UGUAAUCAAGGACUUAUG
Cyclo	42	SEQ. ID 0073	UAAUCAAGGACUUAUGAU
Cyclo	43	SEQ. ID 0074	AUCAAGGACUUAUGAUCC
Cyclo	44	SEQ. ID 0075	CAAGGACUUAUGAUCCAG
Cyclo	45	SEQ. ID 0076	AGGACUUAUGAUCCAGGG
Cyclo	46	SEQ. ID 0077	GACUUAUGAUCCAGGGCG
Cyclo	47	SEQ. ID 0078	CUUAUGAUCCAGGGCGGA
Cyclo	48	SEQ. ID 0079	UCAUGAUCCAGGGCGGAGA
Cyclo	49	SEQ. ID 0080	AUGAUCCAGGGCGGAGACU
Cyclo	50	SEQ. ID 0081	GAUCCAGGGCGGAGACUUC
Cyclo	51	SEQ. ID 0082	UCCAGGGCGGAGACUUCAC
Cyclo	52	SEQ. ID 0083	CAGGGCGGAGACUUCACCA
Cyclo	53	SEQ. ID 0084	GGGCGGAGACUUCACCAGG
Cyclo	54	SEQ. ID 0085	GCGGAGACUUCACCAGGGG
Cyclo	55	SEQ. ID 0086	GGAGACUUCACCAGGGGAG
Cyclo	56	SEQ. ID 0087	AGACUUCACCAGGGGAGAU
Cyclo	57	SEQ. ID 0088	ACUUCACCAGGGGAGAUGG
Cyclo	58	SEQ. ID 0089	UUCACCAGGGGAGAUGGCA
Cyclo	59	SEQ. ID 0090	CACCAGGGGAGAUGGCACA
Cyclo	60	SEQ. ID 0091	CCAGGGGAGAUGGCACAGG

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 4 of 42

Cyclo	61	SEQ. ID 0092	AGGGGAGAUGGCACAGGAG
Cyclo	62	SEQ. ID 0093	GGGAGAUGGCACAGGAGGA
Cyclo	63	SEQ. ID [[0094]] <u>0431</u>	GAGAUGGCACAGGAGGAAA
Cyclo	64	SEQ. ID 0095	GAUGGCACAGGAGGAAAGA
Cyclo	65	SEQ. ID 0094	UGGCACAGGAGGAAAGAGC
Cyclo	66	SEQ. ID 0096	GCACAGGAGGAAAGAGCAU
Cyclo	67	SEQ. ID 0097	ACAGGAGGAAAGAGCAUCU
Cyclo	68	SEQ. ID 0098	AGGAGGAAAGAGCAUCUAC
Cyclo	69	SEQ. ID 0099	GAGGAAAGAGCAUCUACGG
Cyclo	70	SEQ. ID 0100	GGAAAGAGCAUCUACGGUG
Cyclo	71	SEQ. ID 0101	AAAGAGCAUCUACGGUGAG
Cyclo	72	SEQ. ID 0102	AGAGCAUCUACGGUGAGCG
Cyclo	73	SEQ. ID 0103	AGCAUCUACGGUGAGCGCU
Cyclo	74	SEQ. ID 0104	CAUCUACGGUGAGCGCUUC
Cyclo	75	SEQ. ID 0105	UCUACGGUGAGCGCUUCCC
Cyclo	76	SEQ. ID 0106	UACGGUGAGCGCUUCCCCG
Cyclo	77	SEQ. ID 0107	CGGUGAGCGCUUCCCCGAU
Cyclo	78	SEQ. ID 0108	GUGAGCGCUUCCCCGAUGA
Cyclo	79	SEQ. ID 0109	GAGCGCUUCCCCGAUGAGA
Cyclo	80	SEQ. ID 0110	GCGCUUCCCCGAUGAGAAC
Cyclo	81	SEQ. ID 0111	GCUUCCCCGAUGAGAACUU
Cyclo	82	SEQ. ID 0112	UCCCCGAUGAGAACUUCA
Cyclo	83	SEQ. ID 0113	CCCCGAUGAGAACUUCAAA
Cyclo	84	SEQ. ID 0114	CCGAUGAGAACUUCAAACU
Cyclo	85	SEQ. ID 0115	GAUGAGAACUUCAAACUGA
Cyclo	86	SEQ. ID 0116	UGAGAACUUCAAACUGAAG
Cyclo	87	SEQ. ID 0117	AGAACUUCAAACUGAAGCA
Cyclo	88	SEQ. ID 0118	AACUUCAAACUGAAGCACU
Cyclo	89	SEQ. ID 0119	CUUCAAACUGAAGCACUAC
Cyclo	90	SEQ. ID 0120	UCAAACUGAAGCACUACGG
DB	1	SEQ. ID 0121	ACGGGCAAGGCCAAGUGGG
DB	2	SEQ. ID 0122	CGGGCAAGGCCAAGUGGGA
DB	3	SEQ. ID 0123	GGGCAAGGCCAAGUGGGAU

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 5 of 42

DB	4	SEQ. ID 0124	GGCAAGGCCAAGUGGGAUG
DB	5	SEQ. ID 0125	GCAAGGCCAAGUGGGAUGC
DB	6	SEQ. ID 0126	CAAGGCCAAGUGGGAUGCC
DB	7	SEQ. ID 0127	AAGGCCAAGUGGGAUGCCU
DB	8	SEQ. ID 0128	AGGCCAAGUGGGAUGCCUG
DB	9	SEQ. ID 0129	GGCCAAGUGGGAUGCCUGG
DB	10	SEQ. ID 0130	GCCAAGUGGGAUGCCUGGA
DB	11	SEQ. ID 0131	CCAAGUGGGAUGCCUGGAA
DB	12	SEQ. ID 0132	CAAGUGGGAUGCCUGGAAU
DB	13	SEQ. ID 0133	AAGUGGGAUGCCUGGAAUG
DB	14	SEQ. ID 0134	AGUGGGAUGCCUGGAAUGA
DB	15	SEQ. ID 0135	GUGGGAUGCCUGGAAUGAG
DB	16	SEQ. ID 0136	UGGGAUGCCUGGAAUGAGC
DB	17	SEQ. ID 0137	GGGAUGCCUGGAAUGAGCU
DB	18	SEQ. ID 0138	GGAUGCCUGGAAUGAGCUG
DB	19	SEQ. ID 0139	GAUGCCUGGAAUGAGCUGA
DB	20	SEQ. ID 0140	AUGCCUGGAAUGAGCUGAA
DB	21	SEQ. ID 0141	UGCCUGGAAUGAGCUGAAA
DB	22	SEQ. ID 0142	GCCUGGAAUGAGCUGAAAG
DB	23	SEQ. ID 0143	CCUGGAAUGAGCUGAAAGG
DB	24	SEQ. ID 0144	CUGGAAUGAGCUGAAAGGG
DB	25	SEQ. ID 0145	UGGAAUGAGCUGAAAGGGA
DB	26	SEQ. ID 0146	GGAAUGAGCUGAAAGGGAC
DB	27	SEQ. ID 0147	GAAUGAGCUGAAAGGGACU
DB	28	SEQ. ID 0148	AAUGAGCUGAAAGGGACUU
DB	29	SEQ. ID 0149	AUGAGCUGAAAGGGACUUC
DB	30	SEQ. ID 0150	UGAGCUGAAAGGGACUCC
DB	31	SEQ. ID 0151	GAGCUGAAAGGGACUCCA
DB	32	SEQ. ID 0152	AGCUGAAAGGGACUCCAA
DB	33	SEQ. ID 0153	GCUGAAAGGGACUCCAAG
DB	34	SEQ. ID 0154	CUGAAAGGGACUCCAAGG
DB	35	SEQ. ID 0155	UGAAAGGGACUCCAAGGA
DB	36	SEQ. ID 0156	GAAAGGGACUCCAAGGAA

DB	37	SEQ. ID 0157	AAAGGGACUCCAAGGAAG
DB	38	SEQ. ID 0158	AAGGGACUCCAAGGAAGA
DB	39	SEQ. ID 0159	AGGGACUCCAAGGAAGAU
DB	40	SEQ. ID 0160	GGGACUCCAAGGAAGAUG
DB	41	SEQ. ID 0161	GGACUCCAAGGAAGAUGC
DB	42	SEQ. ID 0162	GACUCCAAGGAAGAUGCC
DB	43	SEQ. ID 0163	ACUCCAAGGAAGAUGCCA
DB	44	SEQ. ID 0164	CUCCAAGGAAGAUGCCAU
DB	45	SEQ. ID 0165	UCCAAGGAAGAUGCCAUG
DB	46	SEQ. ID 0166	UCCAAGGAAGAUGCCAUGA
DB	47	SEQ. ID 0167	CCAAGGAAGAUGCCAUGAA
DB	48	SEQ. ID 0168	CAAGGAAGAUGCCAUGAAA
DB	49	SEQ. ID 0169	AAGGAAGAUGCCAUGAAAG
DB	50	SEQ. ID 0170	AGGAAGAUGCCAUGAAAGC
DB	51	SEQ. ID 0171	GGAAGAUGCCAUGAAAGCU
DB	52	SEQ. ID 0172	GAAGAUGCCAUGAAAGCUU
DB	53	SEQ. ID 0173	AAGAUGCCAUGAAAGCUUA
DB	54	SEQ. ID 0174	AGAUGCCAUGAAAGCUUAC
DB	55	SEQ. ID 0175	GAUGCCAUGAAAGCUUACA
DB	56	SEQ. ID 0176	AUGCCAUGAAAGCUUACAU
DB	57	SEQ. ID 0177	UGCCAUGAAAGCUUACAUC
DB	58	SEQ. ID 0178	GCCAUGAAAGCUUACAUCA
DB	59	SEQ. ID 0179	CCAUGAAAGCUUACAUCAA
DB	60	SEQ. ID 0180	CAUGAAAGCUUACAUCAAC
DB	61	SEQ. ID 0181	AUGAAAGCUUACAUCAACA
DB	62	SEQ. ID 0182	UGAAAGCUUACAUCAACAA
DB	63	SEQ. ID 0183	GAAAGCUUACAUCAACAAA
DB	64	SEQ. ID 0184	AAAGCUUACAUCAACAAAG
DB	65	SEQ. ID 0185	AAGCUUACAUCAACAAAGU
DB	66	SEQ. ID 0186	AGCUUACAUCAACAAAGUA
DB	67	SEQ. ID 0187	GCUUACAUCAACAAAGUAG
DB	68	SEQ. ID 0188	CUUACAUCAACAAAGUAGA
DB	69	SEQ. ID 0189	UUACAUCAACAAAGUAGAA

Applicants: Khvorova *et al.*
 Serial No.: 10/714,333
 Filing Date: November 14, 2003
 Preliminary Amendment
 May 20, 2004
 Page 7 of 42

DB	70	SEQ. ID 0190	UACAUCAACAAAGUAGAAG
DB	71	SEQ. ID 0191	ACAUCAACAAAGUAGAAGA
DB	72	SEQ. ID 0192	CAUCAACAAAGUAGAAGAG
DB	73	SEQ. ID 0193	AUCAACAAAGUAGAAGAGC
DB	74	SEQ. ID 0194	UCAACAAAGUAGAAGAGCU
DB	75	SEQ. ID 0195	CAACAAAGUAGAAGAGCUA
DB	76	SEQ. ID 0196	AACAAAGUAGAAGAGCUAA
DB	77	SEQ. ID 0197	ACAAAGUAGAAGAGCUAAA
DB	78	SEQ. ID 0198	CAAAGUAGAAGAGCUAAAG
DB	79	SEQ. ID 0199	AAAGUAGAAGAGCUAAAGA
DB	80	SEQ. ID 0200	AAGUAGAAGAGCUAAAGAA
DB	81	SEQ. ID 0201	AGUAGAAGAGCUAAAGAAA
DB	82	SEQ. ID 0202	GUAGAAGAGCUAAAGAAAA
DB	83	SEQ. ID 0203	UAGAAGAGCUAAAGAAAAA
DB	84	SEQ. ID 0204	AGAAGAGCUAAAGAAAAAA
DB	85	SEQ. ID 0205	GAAGAGCUAAAGAAAAAAU
DB	86	SEQ. ID 0206	AAGAGCUAAAGAAAAAAUA
DB	87	SEQ. ID 0207	AGAGCUAAAGAAAAAAUAC
DB	88	SEQ. ID 0208	GAGCUAAAGAAAAAAUACG
DB	89	SEQ. ID 0209	AGCUAAAGAAAAAAUACGG
DB	90	SEQ. ID 0210	GCUAAAGAAAAAAUACGGG
Luc	1	SEQ. ID 0211	AUCCUCAUAAAGGCCAAGA
Luc	2	SEQ. ID 0212	AGAUCCUCAUAAAGGCCAA
Luc	3	SEQ. ID 0213	AGAGAUCUCAUAAAGGCC
Luc	4	SEQ. ID 0214	AGAGAGAUCCUCAUAAAGG
Luc	5	SEQ. ID 0215	UCAGAGAGAUCCUCAUAAA
Luc	6	SEQ. ID 0216	AAUCAGAGAGAUCCUCAUA
Luc	7	SEQ. ID 0217	AAAAUCAGAGAGAUCCUCA
Luc	8	SEQ. ID 0218	GAAAAUCAGAGAGAUCCU
Luc	9	SEQ. ID 0219	AAGAAAAUCAGAGAGAUCC
Luc	10	SEQ. ID 0220	GCAAGAAAAUCAGAGAGAG
Luc	11	SEQ. ID 0221	ACGCAAGAAAAUCAGAGAG
Luc	12	SEQ. ID 0222	CGACGCAAGAAAAUCAGAG

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 8 of 42

Luc	13	SEQ. ID 0223	CUCGACGCAAGAAAAAUCA
Luc	14	SEQ. ID 0224	AACUCGACGCAAGAAAAAU
Luc	15	SEQ. ID 0225	AAAACUCGACGCAAGAAAA
Luc	16	SEQ. ID 0226	GGAAAACUCGACGCAAGAA
Luc	17	SEQ. ID 0227	CCGGAAAACUCGACGCAAG
Luc	18	SEQ. ID 0228	UACCGGAAAACUCGACGCA
Luc	19	SEQ. ID 0229	CUUACCGGAAAACUCGACG
Luc	20	SEQ. ID 0230	GUCUUACCGGAAAACUCGA
Luc	21	SEQ. ID 0231	AGGUCUUACCGGAAAACUC
Luc	22	SEQ. ID 0232	AAAGGUCUUACCGGAAAAC
Luc	23	SEQ. ID 0233	CGAAAGGUCUUACCGGAAA
Luc	24	SEQ. ID 0234	ACCGAAAGGUCUUACCGGA
Luc	25	SEQ. ID 0235	GUACCGAAAGGUCUUACCG
Luc	26	SEQ. ID 0236	AAGUACCGAAAGGUCUUAC
Luc	27	SEQ. ID 0237	CGAAGUACCGAAAGGUCUU
Luc	28	SEQ. ID 0238	GACGAAGUACCGAAAGGUC
Luc	29	SEQ. ID 0239	UGGACGAAGUACCGAAAGG
Luc	30	SEQ. ID 0240	UGUGGACGAAGUACCGAAA
Luc	31	SEQ. ID 0241	UUUGUGGACGAAGUACCGA
Luc	32	SEQ. ID 0242	UGUUUGUGGACGAAGUACC
Luc	33	SEQ. ID 0243	UGUGUUUGUGGACGAAGUA
Luc	34	SEQ. ID 0244	GUUGUGUUUGUGGACGAAG
Luc	35	SEQ. ID 0245	GAGUUGUGUUUGUGGACGA
Luc	36	SEQ. ID 0246	AGGAGUUGUGUUUGUGGAC
Luc	37	SEQ. ID 0247	GGAGGAGUUGUGUUUGUGG
Luc	38	SEQ. ID 0248	GCGGAGGAGUUGUGUUUGU
Luc	39	SEQ. ID 0249	GCGCGGAGGAGUUGUGUUU
Luc	40	SEQ. ID 0250	UUGCGCGGAGGAGUUGUGU
Luc	41	SEQ. ID 0251	AGUUGCGCGGAGGAGUUGU
Luc	42	SEQ. ID 0252	AAAGUUGCGCGGAGGAGUU
Luc	43	SEQ. ID 0253	AAAAAGUUGCGCGGAGGAG
Luc	44	SEQ. ID 0254	CGAAAAAGUUGCGCGGAGG
Luc	45	SEQ. ID 0255	CGCGAAAAAGUUGCGCGGA

Applicants: Khvorova *et al.*
 Serial No.: 10/714,333
 Filing Date: November 14, 2003
 Preliminary Amendment
 May 20, 2004
 Page 9 of 42

Luc	46	SEQ. ID 0256	ACCGCGAAAAAGUUGCGCG
Luc	47	SEQ. ID 0257	CAACCGCGAAAAAGUUGCG
Luc	48	SEQ. ID 0258	AACAACCGCGAAAAAGUUG
Luc	49	SEQ. ID 0259	GUAACAACCGCGAAAAAGU
Luc	50	SEQ. ID 0260	AAGUAACAACCGCGAAAAA
Luc	51	SEQ. ID 0261	UCAAGUAACAACCGCGAAA
Luc	52	SEQ. ID 0262	AGUCAAGUAACAACCGCGA
Luc	53	SEQ. ID 0263	CCAGUCAAGUAACAACCGC
Luc	54	SEQ. ID 0264	CGCCAGUCAAGUAACAACC
Luc	55	SEQ. ID 0265	GUCGCCAGUCAAGUAACAA
Luc	56	SEQ. ID 0266	ACGUCGCCAGUCAAGUAAC
Luc	57	SEQ. ID 0267	UUACGUCGCCAGUCAAGUA
Luc	58	SEQ. ID 0268	GAUUACGUCGCCAGUCAAG
Luc	59	SEQ. ID 0269	UGGAUUACGUCGCCAGUCA
Luc	60	SEQ. ID 0270	CGUGGAUUACGUCGCCAGU
Luc	61	SEQ. ID 0271	AUCGUGGAUUACGUCGCCA
Luc	62	SEQ. ID 0272	AGAUCGUGGAUUACGUCGC
Luc	63	SEQ. ID 0273	AGAGAUCGUGGAUUACGUC
Luc	64	SEQ. ID 0274	AAAGAGAUCGUGGAUUACG
Luc	65	SEQ. ID 0275	AAAAAGAGAUCGUGGAUUA
Luc	66	SEQ. ID 0276	GGAAAAAGAGAUCGUGGAU
Luc	67	SEQ. ID 0277	ACGGAAAAAGAGAUCGUGG
Luc	68	SEQ. ID 0278	UGACGGAAAAAGAGAUCGU
Luc	69	SEQ. ID 0279	GAUGACGGAAAAAGAGAUC
Luc	70	SEQ. ID 0280	ACGAUGACGGAAAAAGAGA
Luc	71	SEQ. ID 0281	AGACGAUGACGGAAAAAGA
Luc	72	SEQ. ID 0282	AAAGACGAUGACGGAAAAA
Luc	73	SEQ. ID 0283	GGAAAGACGAUGACGGAAA
Luc	74	SEQ. ID 0284	ACGGAAAGACGAUGACGGA
Luc	75	SEQ. ID 0285	GCACGGAAAGACGAUGACG
Luc	76	SEQ. ID 0286	GAGCACGGAAAGACGAUGA
Luc	77	SEQ. ID 0287	UGGAGCACGGAAAGACGAU
Luc	78	SEQ. ID 0288	UUUGGAGCACGGAAAGACG

Luc	79	SEQ. ID 0289	GUUUUGGAGCACGGAAAGA
Luc	80	SEQ. ID 0290	UUGUUUUGGAGCACGGAAA
Luc	81	SEQ. ID 0291	UGUUGUUUUGGAGCACGGA
Luc	82	SEQ. ID 0292	GUUGUUGUUUUGGAGCACG
Luc	83	SEQ. ID 0293	CCGUUGUUGUUUUGGAGCA
Luc	84	SEQ. ID 0294	CGCCGUUGUUGUUUUGGAG
Luc	85	SEQ. ID 0295	GCCGCCGUUGUUGUUUUGG
Luc	86	SEQ. ID 0296	CCGCCGCCGUUGUUGUUUU
Luc	87	SEQ. ID 0297	UCCCGCCGCCGUUGUUGUU
Luc	88	SEQ. ID 0298	CUUCCCGCCGCCGUUGUUG
Luc	89	SEQ. ID 0299	AACUCCCGCCGCCGUUGU
Luc	90	SEQ. ID 0300	UGAACUCCCGCCGCCGUU

Please replace Table V, spanning pages 57-63, with the following amended Table V:

[[**TABLE V**]]

TABLE V					
Gene Name	Accession Number	SEQ. ID NO.	FTIISeqTence	Formula VIII	Formula IX
CLTC	NM_004859	SEQ. ID NO. [[0301]] <u>2400</u>	GAAAGAATCTGTAGAGAAA	76	94.2
CLTC	NM_004859	SEQ. ID NO. [[0302]] <u>2401</u>	GCAATGAGCTGTTTGAAGA	65	39.9
CLTC	NM_004859	SEQ. ID NO. [[0303]] <u>2402</u>	TGACAAAGGTGGATAAATT	57	38.2
CLTC	NM_004859	SEQ. ID NO. [[0304]] <u>2403</u>	GGAAATGGATCTCTTTGAA	54	49.4
CLTA	NM_001833	SEQ. ID NO. [[0305]] <u>2404</u>	GGAAAGTAATGGTCCAACA	22	55.5
CLTA	NM_001833	SEQ. ID NO. [[0306]] <u>2405</u>	AGACAGTTATGCAGCTATT	4	22.9

CLTA	NM_001833	SEQ. ID NO. [[0307]] 2406	CCAATTCTCGGAAGCAAGA	1	17
CLTA	NM_001833	SEQ. ID NO. [[0308]] 2407	GAAAGTAATGGTCCAACAG	-1	-13
CLTB	NM_001834	SEQ. ID NO. [[0309]] 2408	GCGCCAGAGTGAACAAGTA	17	57.5
CLTB	NM_001834	SEQ. ID NO. [[0310]] 2409	GAAGGTGGCCCAGCTATGT	15	-8.6
CLTB	NM_001834	SEQ. ID NO. 0311	GGAACCAGCGCCAGAGTGA	13	40.5
CLTB	NM_001834	SEQ. ID NO. 0312	GAGCGAGATTGCAGGCATA	20	61.7
CALM	U45976	SEQ. ID NO. 0313	GTTAGTATCTGATGACTTG	36	-34.6
CALM	U45976	SEQ. ID NO. 0314	GAAATGGAACCACTAAGAA	33	46.1
CALM	U45976	SEQ. ID NO. 0315	GGAAATGGAACCACTAAGA	30	61.2
CALM	U45976	SEQ. ID NO. 0316	CAACTACACTTTCGAATGC	28	6.8
EPS15	NM_001981	SEQ. ID NO. 0317	CCACCAAGATTTCATGATA	48	25.2
EPS15	NM_001981	SEQ. ID NO. 0318	GATCGGAACTCCAACAAGA	43	49.3
EPS15	NM_001981	SEQ. ID NO. 0319	AAACGGAGCTACAGATTAT	39	11.5
EPS15	NM_001981	SEQ. ID NO. 0320	CCACACAGCATTCTTGTA	33	-23.6
EPS15R	NM_021235	SEQ. ID NO. 0321	GAAGTTACCTTGAGCAATC	48	33
EPS15R	NM_021235	SEQ. ID NO. 0322	GGACTTGCCGATCCAGAA	27	33
EPS15R	NM_021235	SEQ. ID NO. 0323	GCACTTGATCGAGATGAG	20	1.3
EPS15R	NM_021235	SEQ. ID NO. 0324	CAAAGACCAATTCGCGTTA	17	27.7
DNM2	NM_004945	SEQ. ID NO. 0325	CCGAATCAATCGCATCTTC	6	-29.6
DNM2	NM_004945	SEQ. ID NO. 0326	GACATGATCCTGCAGTTCA	5	-14
DNM2	NM_004945	SEQ. ID NO. 0327	GAGCGAATCGTCACCACTT	5	24
DNM2	NM_004945	SEQ. ID NO. 0328	CCTCCGAGCTGGCGTCTAC	-4	-63.6
ARF6	AF93885	SEQ. ID NO. 0329	TCACATGGTTAACCTCTAA	27	-21.1
ARF6	AF93885	SEQ. ID NO. 0330	GATGAGGGACGCCATAATC	7	-38.4
ARF6	AF93885	SEQ. ID NO. 0331	CCTCTAACTACAAATCTTA	4	16.9
ARF6	AF93885	SEQ. ID NO. 0332	GGAAGGTGCTATCCAAAAT	4	11.5
RAB5A	BC001267	SEQ. ID NO. 0333	GCAAGCAAGTCCTAACATT	40	25.1
RAB5A	BC001267	SEQ. ID NO. 0334	GGAAGAGGAGTAGACCTTA	17	50.1
RAB5A	BC001267	SEQ. ID NO. 0335	AGGAATCAGTGTTGTAGTA	16	11.5
RAB5A	BC001267	SEQ. ID NO. 0336	GAAGAGGAGTAGACCTTAC	12	7

RAB5B	NM_002868	SEQ. ID NO. 0337	GAAAGTCAAGCCTGGTATT	14	18.1
RAB5B	NM_002868	SEQ. ID NO. 0338	AAAGTCAAGCCTGGTATTA	6	-17.8
RAB5B	NM_002868	SEQ. ID NO. 0339	GCTATGAACGTGAATGATC	3	-21.1
RAB5B	NM_002868	SEQ. ID NO. 0340	CAAGCCTGGTATTACGTTT	-7	-37.5
RAB5C	AF141304	SEQ. ID NO. 0341	GGAACAAGATCTGTCAATT	38	51.9
RAB5C	AF141304	SEQ. ID NO. 0342	GCAATGAACGTGAACGAAA	29	43.7
RAB5C	AF141304	SEQ. ID NO. 0343	CAATGAACGTGAACGAAAT	18	43.3
RAB5C	AF141304	SEQ. ID NO. 0344	GGACAGGAGCGGTATCACA	6	18.2
EEA1	XM_018197	SEQ. ID NO. 0345	AGACAGAGCTTGAGAATAA	67	64.1
EEA1	XM_018197	SEQ. ID NO. 0346	GAGAAGATCTTTATGCAAA	60	48.7
EEA1	XM_018197	SEQ. ID NO. 0347	GAAGAGAAATCAGCAGATA	58	45.7
EEA1	XM_018197	SEQ. ID NO. 0348	GCAAGTAACTCAACTAACA	56	72.3
AP2B1	NM_001282	SEQ. ID NO. 0349	GAGCTAATCTGCCACATTG	49	-12.4
AP2B1	NM_001282	SEQ. ID NO. 0350	GCAGATGAGTTACTAGAAA	44	48.9
AP2B1	NM_001282	SEQ. ID NO. 0351	CAACTTAATTGTCCAGAAA	41	28.2
AP2B1	NM_001282	SEQ. ID NO. 0352	CAACACAGGATTCTGATAA	33	-5.8
PLK	NM_005030	SEQ. ID NO. 0353	AGATTGTGCTAAGTCTCT	-35	-3.4
PLK	NM_005030	SEQ. ID NO. 0354	ATGAAGATCTGGAGGTGAA	0	-4.3
PLK	NM_005030	SEQ. ID NO. 0355	TTTGAGACTTCTTGCCTAA	-5	-27.7
PLK	NM_005030	SEQ. ID NO. 0356	AGATCACCTCCTTAAATA	15	72.3
GAPDH	NM_002046	SEQ. ID NO. 0357	CAACGGATTTGGTCGTATT	27	-2.8
GAPDH	NM_002046	SEQ. ID NO. 0358	GAAATCCCATCACCATCTT	24	3.9
GAPDH	NM_002046	SEQ. ID NO. 0359	GACCTCAACTACATGGTTT	22	-22.9
GAPDH	NM_002046	SEQ. ID NO. 0360	TGGTTTACATGTTCCAATA	9	9.8
c-Myc		SEQ. ID NO. 0361	GAAGAAATCGATGTTGTTT	31	-11.7
c-Myc		SEQ. ID NO. 0362	ACACAAACTTGAACAGCTA	22	51.3
c-Myc		SEQ. ID NO. 0363	GGAAGAAATCGATGTTGTT	18	26
c-Myc		SEQ. ID NO. 0364	GAAACGACGAGAACAGTTG	18	-8.9
MAP2K1	NM_002755	SEQ. ID NO. 0365	GCACATGGATGGAGGTTCT	26	16
MAP2K1	NM_002755	SEQ. ID NO. 0366	GCAGAGAGAGCAGATTTGA	16	0.4
MAP2K1	NM_002755	SEQ. ID NO. 0367	GAGGTTCTCTGGATCAAGT	14	15.5
MAP2K1	NM_002755	SEQ. ID NO. 0368	GAGCAGATTTGAAGCAACT	14	18.5
MAP2K2	NM_030662	SEQ. ID NO. 0369	CAAAGACGATGACTTCGAA	37	26.4

MAP2K2	NM_030662	SEQ. ID NO. 0370	GATCAGCATTTGCATGGAA	24	-0.7
MAP2K2	NM_030662	SEQ. ID NO. 0371	TCCAGGAGTTTGTCAATAA	17	-4.5
MAP2K2	NM_030662	SEQ. ID NO. 0372	GGAAGCTGATCCACCTTGA	16	59.2
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0373	GCAGAAATCTAAGGATATA	53	35.8
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0374	CAACAAGGATGAAGTCTAT	50	18.3
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0375	CAGCAGAAATCTAAGGATA	41	32.7
KNSL1(EG5)	NM_004523	SEQ. ID NO. 0376	CTAGATGGCTTTCTCAGTA	39	3.9
CyclophilinA_	NM_021130	SEQ. ID NO. 0377	AGACAAGGTCCCAAAGACA	-16	58.1
CyclophilinA_	NM_021130	SEQ. ID NO. 0378	GGAATGGCAAGACCAGCAA	-6	36
CyclophilinA_	NM_021130	SEQ. ID NO. 0379	AGAATTATTCCAGGGTTTA	-3	16.1
CyclophilinA_	NM_021130	SEQ. ID NO. 0380	GCAGACAAGGTCCCAAAGA	8	8.9
LAMIN A/C	NM_170707	SEQ. ID NO. 0381	AGAAGCAGCTTCAGGATGA	31	38.8
LAMIN A/C	NM_170707	SEQ. ID NO. 0382	GAGCTTGACTTCCAGAAGA	33	22.4
LAMIN A/C	NM_170707	SEQ. ID NO. 0383	CCACCGAAGTTCACCCTAA	21	27.5
LAMIN A/C	NM_170707	SEQ. ID NO. 0384	GAGAAGAGCTCCTCCATCA	55	30.1
CyclophilinB	M60857	SEQ. ID NO. 0385	GAAAGAGCATCTACGGTGA	41	83.9
CyclophilinB	M60857	SEQ. ID NO. 0386	GAAAGGATTTGGCTACAAA	53	59.1
CyclophilinB	M60857	SEQ. ID NO. 0387	ACAGCAAATTCATCGTGT	-20	28.8
CyclophilinB	M60857	SEQ. ID NO. 0388	GGAAAGACTGTTCCAAAAA	2	27
DBI1	NM_020548	SEQ. ID NO. 0389	CAACACGCCTCATCCTCTA	27	-7.6
DBI2	NM_020548	SEQ. ID NO. 0390	CATGAAAGCTTACATCAAC	25	-30.8
DBI3	NM_020548	SEQ. ID NO. 0391	AAGATGCCATGAAAGCTTA	17	22
DBI4	NM_020548	SEQ. ID NO. 0392	GCACATACCGCCTGAGTCT	15	3.9
rLUC1		SEQ. ID NO. 0393	GATCAAATCTGAAGAAGGA	57	49.2
rLUC2		SEQ. ID NO. 0394	GCCAAGAAGTTTCCTAATA	50	13.7
rLUC3		SEQ. ID NO. 0395	CAGCATATCTTGAACCATT	41	-2.2
rLUC4		SEQ. ID NO. 0396	GAACAAAGGAAACGGATGA	39	29.2
SeAP1	NM_031313	SEQ. ID NO. 0397	CGGAAACGGTCCAGGCTAT	6	26.9
SeAP2	NM_031313	SEQ. ID NO. 0398	GCTTCGAGCAGACATGATA	4	-11.2
SeAP3	NM_031313	SEQ. ID NO. 0399	CCTACACGGTCCTCCTATA	4	4.9
SeAP4	NM_031313	SEQ. ID NO. 0400	GCCAAGAACCTCATCATCT	1	-9.9
rLUC1		SEQ. ID NO. 0401	GATATGGGCTGAATACAAA	54	40.4
rLUC2		SEQ. ID NO. 0402	GCACTCTGATTGACAAATA	47	54.7

fLUC3		SEQ. ID NO. 0403	TGAAGTCTCTGATTAAGTA	46	34.5
fLUC4		SEQ. ID NO. 0404	TCAGAGAGATCCTCATAAA	40	11.4
mCyclo_1	NM_008907	SEQ. ID NO. 0405	GCAAGAAGATCACCATTTC	52	46.4
mCyclo_2	NM_008907	SEQ. ID NO. 0406	GAGAGAAATTTGAGGATGA	36	70.7
mCyclo_3	NM_008907	SEQ. ID NO. 0407	GAAAGGATTTGGCTATAAG	35	-1.5
mCyclo_4	NM_008907	SEQ. ID NO. 0408	GAAAGAAGGCATGAACATT	27	10.3
BCL2_1	NM_000633	SEQ. ID NO. 0409	GGGAGATAGTGATGAAGTA	21	72
BCL2_2	NM_000633	SEQ. ID NO. 0410	GAAGTACATCCATTATAAG	1	3.3
BCL2_3	NM_000633	SEQ. ID NO. 0411	GTACGACAACCGGGAGATA	1	35.9
BCL2_4	NM_000633	SEQ. ID NO. 0412	AGATAGTGATGAAGTACAT	-12	22.1
BCL2_5	NM_000633	SEQ. ID NO. 0413	TGAAGACTCTGCTCAGTTT	36	19.1
BCL2_6	NM_000633	SEQ. ID NO. 0414	GCATGCGGCCTCTGTTTGA	5	-9.7
QB1	NM_003365.1	SEQ. ID NO. 0415	GCACACAGCUUACUACAUC	52	-4.8
QB2	NM_003365.1	SEQ. ID NO. 0416	GAAAUGCCCUGGUAUCUCA	49	22.1
QB3	NM_003365.1	SEQ. ID NO. 0417	GAAGGAACGUGAUGUGAUC	34	22.9
QB4	NM_003365.1	SEQ. ID NO. 0418	GCACUACUCCUGUGUGUGA	28	20.4
ATE1-1	NM_007041	SEQ. ID NO. 0419	GAACCCAGCUGGAGAACUU	45	15.5
ATE1-2	NM_007041	SEQ. ID NO. 0420	GAUAUACAGUGUGAUCUUA	40	12.2
ATE1-3	NM_007041	SEQ. ID NO. 0421	GUACUACGAUCCUGAUUUAU	37	32.9
ATE1-4	NM_007041	SEQ. ID NO. 0422	GUGCCGACCUUUACAAUUU	35	18.2
EGFR-1	NM_005228	SEQ. ID NO. 0423	GAAGGAAACTGAATTCAAA	68	79.4
EGFR-1	NM_005228	SEQ. ID NO. 0424	GGAAATATGTACTACGAAA	49	49.5
EGFR-1	NM_005228	SEQ. ID NO. 0425	CCACAAAGCAGTGAATTTA	41	7.6
EGFR-1	NM_005228	SEQ. ID NO. 0426	GTAACAAGCTCACGCAGTT	40	25.9

Please replace Table IX, spanning pages 124-129, with the following amended Table IX:

[[Table IX]]

TABLE IX					
Gene Name	Accession #	GI#	Duplex #	Sequence	SEQ. ID NO.
AR	NM_000044	21322251	D-003400-01	GGAAGTCGATCGTATCATT	1369
AR	NM_000044	21322251	D-003400-02	CAAGGGAGGTTACACCAAA	1370
AR	NM_000044	21322251	D-003400-03	TCAAGGAACTCGATCGTAT	1371

AR	NM_000044	21322251	D-003400-04	GAAATGATTGCACTATTGA	1372
ESR1	NM_000125	4503602	D-003401-01	GAATGTGCCTGGCTAGAGA	1373
ESR1	NM_000125	4503602	D-003401-02	CATGAGAGCTGCCAACCTT	1374
ESR1	NM_000125	4503602	D-003401-03	AGAGAAAGATTGGCCAGTA	1375
ESR1	NM_000125	4503602	D-003401-04	CAAGGAGACTCGCTACTGT	1376
ESR2	NM_001437	10835012	D-003402-01	GAACATCTGCTCAACATGA	1377
ESR2	NM_001437	10835012	D-003402-02	GCACGGCTCCATATACATA	1378
ESR2	NM_001437	10835012	D-003402-03	CAAGAAGATTCCCGGCTTT	1379
ESR2	NM_001437	10835012	D-003402-04	GGAAATGCGTAGAAGGAAT	1380
ESRRA	NM_004451	18860919	D-003403-01	GGCCTTCGCTGAGGACTTA	1381
ESRRA	NM_004451	18860919	D-003403-02	TGAATGCACTGGTGTCTCA	1382
ESRRA	NM_004451	18860919	D-003403-03	GCATTGAGCCTCTCTACAT	1383
ESRRA	NM_004451	18860919	D-003403-04	CCAGACAGCGGGCAAAGTG	1384
ESRRB	NM_004452	22035686	D-003404-01	TACCTGAGCTTACAAATTT	1385
ESRRB	NM_004452	22035686	D-003404-02	GCACTTCTATAGCGTCAAA	1386
ESRRB	NM_004452	22035686	D-003404-03	CAACTCCGATTCCATGTAC	1387
ESRRB	NM_004452	22035686	D-003404-04	GGACTCGCCACCCATGTTT	1388
ESRRG	NM_001438	4503604	D-003405-01	AAACAAAGATCGACACATT	1389
ESRRG	NM_001438	4503604	D-003405-02	TCAGGAAACTGTATGATGA	1390
ESRRG	NM_001438	4503604	D-003405-03	GAAGACCAGTCCAAATTAG	1391
ESRRG	NM_001438	4503604	D-003405-04	ATGAAGCGCTGCAGGATTA	1392
HNF4A	NM_000457	21361184	D-003406-01	CGACATCACTGGAGCATAT	1393
HNF4A	NM_000457	21361184	D-003406-02	GAAGGAAGCCGTCCAGAAT	1394
HNF4A	NM_000457	21361184	D-003406-03	CCAAGTACATCCCAGCTTT	1395
HNF4A	NM_000457	21361184	D-003406-04	GGACATGGCCGACTACAGT	1396
HNF4G	NM_004133	6631087	D-003407-01	GCACTGACATAAACGTTAA	1397
HNF4G	NM_004133	6631087	D-003407-02	ACAAAGAGATCCATGATGT	1398
HNF4G	NM_004133	6631087	D-003407-03	AGAGATCCATGATGTATAA	1399
HNF4G	NM_004133	6631087	D-003407-04	AAATGAACGTGACAGAATA	1400
HSAJ2425	NM_017532	8923776	D-003408-01	GAATGAATCTACACCTTTG	1401
HSAJ2425	NM_017532	8923776	D-003408-02	GGAAATACGTGGAGACACT	1402
HSAJ2425	NM_017532	8923776	D-003408-03	CCAGATAACTACGGCGATA	1403
HSAJ2425	NM_017532	8923776	D-003408-04	TGGCGTACCTTCTCATTGA	1404
NR0B1	NM_000475	5016089	D-003409-01	CAGCATGGATGATATGATG	1405
NR0B1	NM_000475	5016089	D-003409-02	CTGCTGAGATTCATCAATG	1406
NR0B1	NM_000475	5016089	D-003409-03	ACAGATTCATCGAACTTAA	1407
NR0B1	NM_000475	5016089	D-003409-04	GAACGTGGCGCTCCTGTAC	1408
NR0B2	NM_021969	13259502	D-003410-01	GAATATGCCTGCCTGAAAG	1409
NR0B2	NM_021969	13259502	D-003410-02	GGAATATGCCTGCCTGAAA	1410

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 16 of 42

NR0B2	NM_021969	13259502	D-003410-03	CGTAGCCGCTGCCTATGTA	1411
NR0B2	NM_021969	13259502	D-003410-04	GCCATTCTCTACGCACTTC	1412
NR1D1	NM_021724	13430847	D-003411-01	CAACACAGGTGGCGTCATC[[T T]]	1413
NR1D1	NM_021724	13430847	D-003411-02	GGCATGGTGTTACTGTGTA[[TT]]	1414
NR1D1	NM_021724	13430847	D-003411-03	CAACATGCATTCCGAGAAG[[T T]]	1415
NR1D1	NM_021724	13430847	D-003411-04	GCGCTTTGCTTCGTTGTTC[[TT]]	1416
NR1H2	NM_007121	11321629	D-003412-01	GAACAGATCCGGAAGAAGA	1417
NR1H2	NM_007121	11321629	D-003412-02	GAAGAACAGATCCGGAAGA	1418
NR1H2	NM_007121	11321629	D-003412-03	CTAAGCAAGTGCCTGGTTT	1419
NR1H2	NM_007121	11321629	D-003412-04	GCTAACAGCGGCTCAAGAA	1420
NR1H3	NM_005693	5031892	D-003413-01	GAACAGATCCGCCTGAAGA	1421
NR1H3	NM_005693	5031892	D-003413-02	GGAGATAGTTGACTTTGCT	1422
NR1H3	NM_005693	5031892	D-003413-03	GAGTTTGCCTTGCTCATTG	1423
NR1H3	NM_005693	5031892	D-003413-04	TGACTTTGCTAAACAGCTA	1424
NR1H4	NM_005123	4826979	D-003414-01	CAAGTGACCTCGACAACAA	1425
NR1H4	NM_005123	4826979	D-003414-02	GAAAGAATTCGAAATAGTG	1426
NR1H4	NM_005123	4826979	D-003414-03	CAACAGACTCTTCTACATT	1427
NR1H4	NM_005123	4826979	D-003414-04	GAACCATACTCGCAATACA	1428
NR1I2	NM_003889	11863133	D-003415-01	GAACCATGCTGACTTTGTA	1429
NR1I2	NM_003889	11863133	D-003415-02	GATGGACGCTCAGATGAAA	1430
NR1I2	NM_003889	11863133	D-003415-03	CAACCTACATGTTCAAAGG	1431
NR1I2	NM_003889	11863133	D-003415-04	CAGGAGCAATTCGCCATTA	1432
NR1I3	NM_005122	4826660	D-003416-01	GGAAATCTGTCACATCGTA	1433
NR1I3	NM_005122	4826660	D-003416-02	TCGCAGACATCAACACTTT	1434
NR1I3	NM_005122	4826660	D-003416-03	CCTCTTCGCTACACAATTG	1435
NR1I3	NM_005122	4826660	D-003416-04	GAACAGTTTGTGCAGTTTA	1436
NR2C1	NM_003297	4507672	D-003417-01	TGACAGCACTTGATCATAA	1437
NR2C1	NM_003297	4507672	D-003417-02	GGAAGGAAGTGTACACCTA	1438
NR2C1	NM_003297	4507672	D-003417-03	GAGCACATCTTCAAACCTAC	1439
NR2C1	NM_003297	4507672	D-003417-04	GAAGAAATTGCACATCAAA	1440
NR2C2	NM_003298	4507674	D-003418-01	GAACAACGGTGACACTTCA	1441
NR2C2	NM_003298	4507674	D-003418-02	CTGATGAGCTCCAACATAA	1442
NR2C2	NM_003298	4507674	D-003418-03	CAACCTAAGTGAATCTTTG	1443
NR2C2	NM_003298	4507674	D-003418-04	GAAGACACCTACCGATTGG	1444
NR2E1	NM_003269	21361108	D-003419-01	GATCATATCTGAAATACAG	1445
NR2E1	NM_003269	21361108	D-003419-02	CAAGACTGCTTTCAGATAT	1446

NR2E1	NM_003269	21361108	D-003419-03	GTTAGATGCTACTGAATTT	1447
NR2E1	NM_003269	21361108	D-003419-04	CAATGTATCTCTATGAAGT	1448
NR2E3	NM_014249	7657394	D-003420-01	GAGAAGCTCCTTTGTGATA	1449
NR2E3	NM_014249	7657394	D-003420-02	GAAGCACTATGGCATCTAT	1450
NR2E3	NM_014249	7657394	D-003420-03	GAAGGATCCTGAGCACGTA	1451
NR2E3	NM_014249	7657394	D-003420-04	GAAGCTCCTTTGTGATATG	1452
NR2F1	NM_005654	20127484	D-003421-01	GAAACTCTCATCCGCGATA	1453
NR2F1	NM_005654	20127484	D-003421-02	TCTCATCCGCGATATGTTA	1454
NR2F1	NM_005654	20127484	D-003421-03	CAAGAAGTGCCTCAAAGTG	1455
NR2F1	NM_005654	20127484	D-003421-04	GGAACTTAACTTACACATG	1456
NR2F2	NM_021005	14149745	D-003422-01	GTACCTGTCCGGATATATT	1457
NR2F2	NM_021005	14149745	D-003422-02	CCAACCAGCCGACGAGATT	1458
NR2F2	NM_021005	14149745	D-003422-03	ACTCGTACCTGTCCGGATA	1459
NR2F2	NM_021005	14149745	D-003422-04	GGCCGTATATGGCAATTCA	1460
NR2F6	NM_005234	20070198	D-003423-01	CGACGCCTGTGGCCTCTCA	1461
NR2F6	NM_005234	20070198	D-003423-02	CAGCCGGTGTCCGAAGTGA	1462
NR2F6	NM_005234	20070198	D-003423-03	CAACCGTGACTGCCAGATC	1463
NR2F6	NM_005234	20070198	D-003423-04	GTAAGTCCGTCTCAAGAAG	1464
NR3C1	NM_000176	4504132	D-003424-01	GAGGACAGATGTACCACTA	1465
NR3C1	NM_000176	4504132	D-003424-02	GATAAGACCATGAGTATTG	1466
NR3C1	NM_000176	4504132	D-003424-03	GAAGACGATTCATTCTTT	1467
NR3C1	NM_000176	4504132	D-003424-04	GGACAGATGTACCACTATG	1468
NR3C2	NM_000901	4505198	D-003425-01	GCAAACAGATGATCCAAGT	1469
NR3C2	NM_000901	4505198	D-003425-02	CAGCTAAGATTTATCAGAA	1470
NR3C2	NM_000901	4505198	D-003425-03	GCACGAAAGTCAAAGAAGT	1471
NR3C2	NM_000901	4505198	D-003425-04	GGTATCCGGTCTTAGAATA	1472
NR4A1	NM_002135	21361341	D-003426-01	GAAGGAAGTTGTCCGAACA	1473
NR4A1	NM_002135	21361341	D-003426-02	CAGGAGAGTTTGACACCTT	1474
NR4A1	NM_002135	21361341	D-003426-03	CAGTGGCTCTGACTACTAT	1475
NR4A1	NM_002135	21361341	D-003426-04	GAAGGCCGCTGTGCTGTGT	1476
NR4A2	NM_006186	5453821	D-003427-01	GCAATGCGTTTCGTGGCTTT	1477
NR4A2	NM_006186	5453821	D-003427-02	CGGCTACACAGGAGAGTTT	1478
NR4A2	NM_006186	5453821	D-003427-03	CCACGTGACTTTCAACAAT	1479
NR4A2	NM_006186	5453821	D-003427-04	GAATACAGCTCCGATTCT	1480
NR4A3	NM_006981	11276070	D-003428-01	CAAAGAAGATCAGACATTA	1481
NR4A3	NM_006981	11276070	D-003428-02	GATCAGACATTACTTATTG	1482
NR4A3	NM_006981	11276070	D-003428-03	CCAGAGATCTTGATTATTC	1483
NR4A3	NM_006981	11276070	D-003428-04	GAAGTTGTCCGTACAGATA	1484
NR5A1	NM_004959	20070192	D-003429-01	GATTTGAAGTTCCTGAATA	1485

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 18 of 42

NR5A1	NM_004959	20070192	D-003429-02	GGAGCGAGCTGCTGGTGTT	1486
NR5A1	NM_004959	20070192	D-003429-03	GGAGGTGGCCGACCAGATG	1487
NR5A1	NM_004959	20070192	D-003429-04	CAACGTGCCTGAGCTCATC	1488
NR5A2	NM_003822	20070161	D-003430-01	CCAAACATATGGCCACTTT	1489
NR5A2	NM_003822	20070161	D-003430-02	TCAGAGAACTTAAGGTTGA	1490
NR5A2	NM_003822	20070161	D-003430-03	GGATCCATCTTCCTGGTTA	1491
NR5A2	NM_003822	20070161	D-003430-04	AAGAATACCTCTACTACAA	1492
NR6A1	NM_033334	15451847	D-003431-01	CAACGAACCTGTCTCATTT	1493
NR6A1	NM_033334	15451847	D-003431-02	GAAGAACTACACAGATTTA	1494
NR6A1	NM_033334	15451847	D-003431-03	GAAGATGGATACGCTGTGA	1495
NR6A1	NM_033334	15451847	D-003431-04	AAACGATACTGGTACATTT	1496
null	D16815	2116671	D-003432-01	GAAGAATGATCGAATAGAT	1497
null	D16815	2116671	D-003432-02	GAACATGGAGCAATATAAT	1498
null	D16815	2116671	D-003432-03	GAGGAGCTCTTGGCCTTTA	1499
null	D16815	2116671	D-003432-04	TAAACAACATGCACTCTGA	1500
PGR	NM_000926	4505766	D-003433-01	GAGATGAGGTCAAGCTACA	1501
PGR	NM_000926	4505766	D-003433-02	CAGCGTTTCTATCAACTTA	1502
PGR	NM_000926	4505766	D-003433-03	AGATAACTCTCATTGAGTA	1503
PGR	NM_000926	4505766	D-003433-04	GTAGTCAAGTGGTCTAAAT	1504
PPARA	NM_005036	7549810	D-003434-01	TCACGGAGCTCACGGAATT	1505
PPARA	NM_005036	7549810	D-003434-02	GAACATGACATAGAAGATT	1506
PPARA	NM_005036	7549810	D-003434-03	GGATAGTTCTGGAAGCTTT	1507
PPARA	NM_005036	7549810	D-003434-04	GACTCAAGCTGGTGTATGA	1508
PPARD	NM_006238	5453939	D-003435-01	GAGCGCAGCTGCAAGATTC	1509
PPARD	NM_006238	5453939	D-003435-02	GCATGAAGCTGGAGTACGA	1510
PPARD	NM_006238	5453939	D-003435-03	GGAAGCAGTTGGTGAATGG	1511
PPARD	NM_006238	5453939	D-003435-04	GCTGCAAGATTCAGAAGAA	1512
PPARG	NM_138712	20336234	D-003436-01	AGACTCAGCTCTACAATAA	1513
PPARG	NM_138712	20336234	D-003436-02	GATTGAAGCTTATCTATGA	1514
PPARG	NM_138712	20336234	D-003436-03	AAGTAACTCTCCTCAAATA	1515
PPARG	NM_138712	20336234	D-003436-04	GCATTTCTACTCCACATTA	1516
RARA	NM_000964	4506418	D-003437-01	GACAAGAACTGCATCATCA	1517
RARA	NM_000964	4506418	D-003437-02	GCAAATACACTACGAACAA	1518
RARA	NM_000964	4506418	D-003437-03	GAACAACAGCTCAGAACAA	1519
RARA	NM_000964	4506418	D-003437-04	GAGCAGCAGTTCTGAAGAG	1520
RARB	NM_000965	14916493	D-003438-01	GCACACTGCTCAATCAATT	1521
RARB	NM_000965	14916493	D-003438-02	GCAGAAGTATTCAGAAGAA	1522
RARB	NM_000965	14916493	D-003438-03	GGAATGACAGGAACAAGAA	1523
RARB	NM_000965	14916493	D-003438-04	GCACAGTCCTAGCATCTCA	1524

RARG	NM_000966	21359851	D-003439-01	GAAATGACCGGAACAAGAA	1525
RARG	NM_000966	21359851	D-003439-02	TAGAAGAGCTCATCACCAA	1526
RARG	NM_000966	21359851	D-003439-03	CAAGGAAGCTGTGCGAAAT	1527
RARG	NM_000966	21359851	D-003439-04	TCAGTGAGCTGGCTACCAA	1528
RORA	NM_134261	19743902	D-003440-01	GGAAAGAGTTTATGTTCTA	1529
RORA	NM_134261	19743902	D-003440-02	CAAGATCTGTGGAGACAAA	1530
RORA	NM_134261	19743902	D-003440-03	GCACCTGACTGAAGATGAA	1531
RORA	NM_134261	19743902	D-003440-04	CCGAGAAGATGGAATACTA	1532
RORB	NM_006914	19743906	D-003441-01	GCACAGAACATCATTAAAGT	1533
RORB	NM_006914	19743906	D-003441-02	CCACACCTATGAAGAAATT	1534
RORB	NM_006914	19743906	D-003441-03	GATCAAATTCTACTTCTGA	1535
RORB	NM_006914	19743906	D-003441-04	TCAAACAGATAAAGCAAGA	1536
RORC	NM_005060	19743908	D-003442-01	TAGAACAGCTGCAGTACAA	1537
RORC	NM_005060	19743908	D-003442-02	TCACCGAGGCCATTGAGTA	1538
RORC	NM_005060	19743908	D-003442-03	GAACAGCTGCAGTACAATC	1539
RORC	NM_005060	19743908	D-003442-04	CCTCATGCCACCTTGAATA	1540
RXRA	NM_002957	21536318	D-003443-01	TGACGGAGCTTGTGTCCAA	1541
RXRA	NM_002957	21536318	D-003443-02	CAACAAGGACTGCCTGATT	1542
RXRA	NM_002957	21536318	D-003443-03	GCAAGGACCTGACCTACAC	1543
RXRA	NM_002957	21536318	D-003443-04	GCAAGGACCGGAACGAGAA	1544
RXRB	NM_021976	21687229	D-003444-01	GCAAAGACCTTACATACTC	1545
RXRB	NM_021976	21687229	D-003444-02	GCAATCATTCTGTTTAATC	1546
RXRB	NM_021976	21687229	D-003444-03	TCACACCGATCCATTGATG	1547
RXRB	NM_021976	21687229	D-003444-04	GCAAACGGCTATGTGCAAT	1548
RXRG	NM_006917	21361386	D-003445-01	GGAAGGACCTCATCTACAC	1549
RXRG	NM_006917	21361386	D-003445-02	CCGGATCTCTGGTTAAACA	1550
RXRG	NM_006917	21361386	D-003445-03	GCGAGCCATTGTACTCTTT	1551
RXRG	NM_006917	21361386	D-003445-04	GAGCCATTGTACTCTTTAA	1552
THRA	NM_003250	20127451	D-003446-01	GGACAAAGACGAGCAGTGT	1553
THRA	NM_003250	20127451	D-003446-02	GGAAACAGAGGCGGAAATT	1554
THRA	NM_003250	20127451	D-003446-03	GTAAGCTGATTGAGCAGAA	1555
THRA	NM_003250	20127451	D-003446-04	GAACCTCCATCCCACCTAT	1556
THRB	NM_000461	10835122	D-003447-01	GAATGTCGCTTTAAGAAAT	1557
THRB	NM_000461	10835122	D-003447-02	GAACAGTCGTCGCCACATC	1558
THRB	NM_000461	10835122	D-003447-03	GGACAAGCACCAATAGTCA	1559
THRB	NM_000461	10835122	D-003447-04	GTGGAAAGGTTGACTTGGA	1560
VDR	NM_000376	4507882	D-003448-01	TGAAGAAGCTGAACTTGCA	1561
VDR	NM_000376	4507882	D-003448-02	GCAACCAAGACTACAAGTA	1562
VDR	NM_000376	4507882	D-003448-03	TCAATGCTATGACCTGTGA	1563
VDR	NM_000376	4507882	D-003448-04	CCATTGAGGTCATCATGTT	1564

Please replace Table X, spanning pages 129-146, with the following amended Table X:

Table X

TABLE X		
Gene Symbol	Sense	SEQ ID NO
ABCB1	GACCAUAAAUGUAAGGUUU	1565
	UAGAAGAUCUGAUGUCAAA	1566
	GAAAGUUCACUUCAGUUA	1567
	GAAGAUCGCUACUGAAGCA	1568
ABCC1	Sense	
	GGAAGCAACUGCAGAGACA	1569
	GAUGACACCUCUCAACAAA	1570
	UAAAGUUGCUCAUCAAGUU	1571
ABCG2	Sense	
	GCAGAUGCCUUCUUCGUUA	1573
	AGGCAAAUCUUCGUUAUUA	1574
	GGAAGAAAUCUGGUCUAA	1575
KCNH2	UGACUCAUCCCAACAUUUA	1576
	Sense	
	CCGACGUGCUGCCUGAGUA	1577
	GAGAAGAGCAGCGACACUU	1578
KCNH1	GAUCAUAGCACCUAAGUA	1579
	GCUAUUUACUGCUCUUAUU	1580
	UCACUGGGCUCCUUUAAUU	1581
	GUGCGAGCCUUCUGAAUUAU	1582
CLCA1	GCUAAGCUAUACUACUGUA	1583
	UGACGGCGCUCUACUUCAC	1584
	Sense	
	GAGAUGAAUCCUUUGAAA	1585
CLCA1	GAAGAACGCAUGAAACGAA	1586
	GAUAAAGACACGAUUGAAA	1587
	GCUGAGAGGUCUAUUUAAA	1588
	Sense	
	GAACAACAAUGGCUAUGAA	1589
	GUACAUACCUGGCUGGAUU	1590

Applicants: Khvorova *et al.*
 Serial No.: 10/714,333
 Filing Date: November 14, 2003
 Preliminary Amendment
 May 20, 2004
 Page 21 of 42

	GAACAGCUCACAAGUAUUAU	1591
	GGAAACGUGUGUCUAUAUU	1592
SLC6A1	Sense	
	GGAGGUGGGAGGACAGUUA	1593
	UCACAGCCCUGGUGGAUGA	1594
	GAAGCUGGCUCUUAUGUUC	1595
	GGUCAACACUACCAACAUG	1596
SLC6A2	Sense	
	GAACACAAGGUCAACAUUG	1597
	AGAAGGAGCUGGCCUAGUG	1598
	CGGAAACUCUUCACAUUUG	1599
	CAACAAAUUUGACAACAAC	1600
SLC21A2	Sense	
	GUACAUCUCCAUCUUAUUU	1601
	GGAAGUGGCUGAGUUAUUA	1602
	GAAGGGAGGCUCAAUGUAA	1603
	GAAGGAAGUGGCUGAGUUA	1604
SLC21A3	Sense	
	GUAGAAACAGGAGCUAUUA	1605
	CAAGAUUACUGUCAAAACAA	1606
	GCACAAGAGUAUUUGGUAA	1607
	GCAAAUGUCCCUUCUGUAU	1608
	GCAUGACUCCUAUAUAAUA	1609
	AAACAGCAAUUUCCCUUAA	1610
	GAAAAUGCCUCUUCAGGAA	1611
SLC28A1	Sense	
	GUUCAUCGCUCUCCUCUUU	1612
	GGAUCAAGCUGUUUCUGAA	1613
	GGACUGCAGUUUGUACUUG	1614
	GAGUGAAACUGACCUAUGG	1615
SLC29A1	Sense	
	GAACGCUGCUCUCCGUGGAA	1616
	GAAAGCCACUCUAUCAAG	1617
	GAAACCAGGUGCCUUCAGA	1618
	CCUCACAGCUGUAUUCAUG	1619
SLC26A1	Sense	
	CCACGGAGCUGCUGGUCAU	1620
	GGGUUGACAUCUUAUUUGA	1621
	GCACGAGGGUCUCUGUGUU	1622
	GGCCAUCGCCUACUCAUUG	1623
	CAACACCCAUGGCAAUUA	1624
	GAGGAAAGAUCUUGCUGAU	1625

SLC26A2	GAGCAAGCGUCCUCCAAAU	1626
	GCAACACCCAUGGCAAUUA	1627
	Sense	
	CCAAAGAACUCAUGAACA	1628
	ACAAGAACCUUCAGACUAA	1629
SLC4A4	GAAGGUAGAUAGAAGAAUG	1630
	GUUUUGAACUGUACUGUAA	1631
	Sense	
	GCAAUUCUCUUCUUUAUC	1632
	GGAAAGAUGUCCACUGAAA	1633
GLRA1	GGACAAAGCCUUCUCAAU	1634
	GGAAUGGGAUCCAGCAAUU	1635
	Sense	
	UGAAAGCCAUUGACAUUUG	1636
	CAGACACGCUGGAGUUUAA	1637
KLK1	CAAUAGCGCUUUCUGGUUU	1638
	GCAGGUAGCAGAUGGACUA	1639
	Sense	
	UCAGAGUGCUGUCUUAUGU	1640
	CAACUUGUUUGACGACGAA	1641
ADAM2	UGACAGAGCCUGCUGAUAC	1642
	AGGCGGCUCUGUACCAUUU	1643
	Sense	
	GAAACAUGCUGUGAUUUG	1644
	GCAGAUGUUUCCUUAUUA	1645
XPNPEP1	CAACAGAGAUGCCAUGAUA	1646
	GAAAGGCGCUACAUUGAGA	1647
	Sense	
	GACCUGAGCUUCCCAACAA	1648
	GCGACUGGCUCAACAAUUA	1649
GZMA	GAGAUUGCGUGGCUAUUUA	1650
	GACAGCAACUGGACACUUA	1651
	Sense	
	GGAAGAGACUCGUGCAAUG	1652
	GGAACCAUGUGCCAAGUUG	1653
CMKLR1	GAAGUAACUCCUCAUUCAA	1654
	GAACUCCUAUAGAUUUCUG	1655
	Sense	
	CAUAGAAGCUUUACCAAGA	1656
	GAAUGGAGGAUGAAGAUUA	1657

	GGUCAAUGCUCUAAGUGAA	1658
	GAGAGGACUUCUAUGAAUG	1659
CLN3	Sense	
	CAUCAUGCCUUCUGAAUAA	1660
	CAACAGCUCaucACGAUUU	1661
	GCAACAACUUCUCUUAUGU	1662
	GGUCUUCGCUAGCAUCUCA	1663
CALCR	Sense	
	GGACCUAGCUGUUGUAAAG	1664
	GAAAGACCAUGCAUUUAAA	1665
	GCAGGAAGAUGUAUGCUUU	1666
	GAAUAAACCAGUAUCGUUA	1667
OXTR	Sense	
	GGACCCAGAUAUCCAAUA	1668
	GCAAUACUAUCCUAACUGA	1669
	GAAUAUAGAUUAGCGUUUG	1670
	GAUGAGGCAUGACUACUAA	1671
EDG4	Sense	
	GCGAGUCUGUCCACUAUAC	1672
	GAGAACGGCCACCCACUGA	1673
	GAACGGCCACCCACUGAUG	1674
	GGUCAAUGCUGCUGUGUAC	1675
EDG5	Sense	
	UCCAGGAACACUAUAUAUA	1676
	GUGACCAUCUUCUCCAUA	1677
	CAUCCUCUGUUGCGCCAUU	1678
	CCAACAAGGUCCAGGAACA	1679
EDG7	Sense	
	ACACUGAUACUGUCGAUGA	1680
	AAUAGGAGCAACACUGAUA	1681
	CAGCAGGAGUUACCUUGUU	1682
	GGACACCCAUGAAGCUAAU	1683
PTCH	Sense	
	GCACAGAACUCCACUCAA	1684
	GGACAGCAGUUCAUUGUUA	1685
	GAGAAGAGGCUAUGUUUAA	1686
	GGACAAACUUCGACCCUUU	1687
SMO	Sense	
	UCGCUACCCUGCUGUUAUU	1688
	GCUACAAGAACUACCGAUA	1689
	CAAGAAAGCUUCCUUAAC	1690

	GAGAAGAAAUACAGUCAAU	1691
CASP3	Sense	
	CAAUUAUUCUGAAGAGCUA	1692
	GAACUGGACUGUGGCAUUG	1693
	GUGAGAAGAUGGUUAUUU	1694
	GAGGGUACUUUAAGACAU	1695
CASP6	Sense	
	CAUGAGGUGUCAACUGUUA	1696
	GAAGUGAAAUGCUUUAUG	1697
	AAUAUGGCCUCCUCCUAG	1698
	GCAAUCAAUUUAUGCAUA	1699
	CAACAUAAACUGAGGUGGAU	1700
	CAUGGUACAUUCAAGAUUU	1701
CASP7	Sense	
	GAACUCUACUUCAGUCAAU	1704
	GGGCAAAUGCAUCAUAAUA	1703
	CAACAGAGGGAGUUUAAUA	1704
	GAACAAAGCCACUGACUGA	1705
CASP8	Sense	
	GAAGUGAACUAUGAAGUAA	1706
	CAACAAGGAUGACAAGAAA	1707
	GGACAAAGUUUACCAAUG	1708
	GAGGGUCGAUCAUCUAUUA	1709
	GAAUAUAGAGGGCUUAUGA	1710
	CAACGACUAUGAAGAAUUC	1711
	GAAGUGAGCAGAUCAAGAU	1712
	GAGGAAAUCUCCAAAUGCA	1713
CASP9	Sense	
	CCAGGCAGCUGAUCAUAGA	1714
	UCUCAGGUGUUGCCAAUA	1715
	GAACAGCUGUAAUCUAUGA	1716
	CCACUGGUCUGUAGGGAUU	1717
DVL1	Sense	
	UCGUAAAGCUGUUGAUUAUC	1718
	GAGGAGAUCUUUGAUGACA	1719
	GUAAAGCUGUUGAUUAUCGA	1720
	GAUCGUAAAGCUGUUGAUA	1721
DVL2	Sense	
	AGACGAAGGUGAUUUACCA	1722
	UGUGAGAGCUACCUAGUCA	1723
	GAAGAAAUUUCAGAUGACA	1724
	UAAUAGGCAUUUCCUCUUU	1725

PTEN	Sense	
	GUGAAGAUCUUGACCAAUG	1726
	GAUCAGCAUACACAAUUA	1727
	GAAUGAACCUUCUGCAACA	1728
	GGCGCUAUGUGUAUUAUA	1729
PDK1	Sense	
	GUACAAAGCUGGUUAUUAUCC	1730
	GAAAGACUCCCAGUGUAUA	1731
	GGAAGUCCAUCUCAUCGAA	1732
	CCAAAGACAUGACGACGUU	1733
PDK2	Sense	
	GUAAAGAGGAGACUGAAUG	1734
	GGUCUGUGAUGGUCCCUAA	1735
	CAAAGAUGCCUACGACAUG	1736
	GGGCGAUGCCUGAGGGUUA	1737
PPP2CA	Sense	
	UCACACAAGUUUAUGGUUU	1738
	CAACAGCCGUGACCACUUU	1739
	UAACCAAGCUGCAAUCAUG	1740
	GAACUUGACGAUACUCUAA	1741
CTNNA1	Sense	
	GAAGAGAGGUCGUUCUAAG	1742
	AAGCAGAUGUGCAUGAUUA	1743
	UCUAAUAACUGCAGUGUUU	1744
	GUAAAGGGCCCUCUAAUAA	1745
CTNNA2	Sense	
	GAAAGAAUAUGCCCAAGUU	1746
	GAAGAAGAAUGCCACAAUG	1747
	GCAGGAAGAUUAUGAUGUG	1748
	AAAGAAAGCCCAUGUACUA	1749
HSPCA	Sense	
	GGGAAAGAGCUGCAUAUUA	1750
	GCUUAGAACUCUUUACUGA	1751
	UAUAAGAGCUUGACCAAUG	1752
	GCAGAUUAUCUCUAUGAUUG	1753
DCTN2	Sense	
	CAACUCAUGUCCAAUACUG	1754
	GGAAUGAGCCAGAUGUUUA	1755
	GGAGACAGCUGUACGUUGU	1756
	UCCAAGAGCUGACAACUGA	1757

CD2	Sense	
	GUAAGGAGAAGCAAUAUAA	1758
	AAGAUGAGCUUCCAUGUA	1759
	GGACAUCUAUCUCAUUAU	1760
	GACAAGAGCCCACAGAGUA	1761
BAD	Sense	
	GUACUUCCCUCAGGCCUAU	1762
	GCUGUGCCUUGACUACGUA	1763
	GUACUUCCCUCAGGCCUAU	1764
	GGUCAGGUGCCUCGAGAUC	1765
SMAC	Sense	
	CAGCGUAACUUCUUCUUC	1766
	UAACUUCUUCUUCAGGUA	1767
	CAGCUGCUCUUACCCAUUU	1768
	GAUUGAAGCUAUUACUGAA	1769
	UAGAAGAGCUCGUCAGAA	1770
	CCACAUUAUGCGUUGAUUGA	1771
	GCGCAGGGCUCUCUACCUA	1772
MAP3K5	Sense	
	GAACAGCCUUCAAAUCAAA	1773
	GAUGUUCUCUACUAUGUUA	1774
	GCAAAUACUGGAAGGAUUA	1775
PVR	Sense	
	CCACACGGCUGACCUCUAU	1777
	CAGCAGAAUCCUCUUAUA	1778
	GCAGAAUCCUCUUAUAAA	1779
ERBB2	Sense	
	UGUGGGAGCUGAUGACUUU	1781
	UCACAGAGAUCUUGAAAGG	1782
	UGGAAGAGAUACAGGUUA	1783
SOS1	Sense	
	GAGCACCACUUCUAUGAUU	1785
	CAAAGAAGCUGUUCAAUAU	1786
	UGAAAGCCCUCUUAUUA	1787
	GAAAUAGCAUGGAGAAGGA	1788
BRCA1	Sense	
	CCAUACAGCUUCAUAAUA	1789
	GAAGAGAACUUAUCUAGUG	1790
	GAAGUGGGCUCAGUAUUA	1791

	GCAAGAUGCUGAUUCAUUA	1792
	GAAGUGGGCUCCAGUAUUA	1793
	GAACGGACACUGAAAUAUU	1794
	GCAGAUAGUUCUACCAGUA	1795
CDKN1A	Sense	
	GAACAAGGAGUCAGACAUU	1796
	AAACUAGGCGGUUGAAUGA	1797
	GAUGGAACUUCGACUUUGU	1798
	GUAAACAGAUGGCACUUUG	1799
CDKN1B	Sense	
	GGAAUGGACAUCUGUAUA	1800
	GGAGAAAGAUGUCAAAACGU	1801
	GAAUGGACAUCUGUAUAA	1802
	GUAAACAGCUCGAAUUAAG	1803
SLC2A4	Sense	
	CAGAUAGGCUCCGAAGAUG	1804
	AGACUCAGCUCAGAAUAC	1805
	GAUCGGUUCUUUCAUCUUC	1806
	CAGGAUCGGUUCUUUCAUC	1807
NOS2A	Sense	
	CCAGAUAAAGUGACAUAAAGU	1808
	UAAGUGACCUGCUUUGUAA	1809
	GAAGAGAGAUUCCAUGAA	1810
	UGAAAGAGCUCACAACAA	1811
FRAP1	Sense	
	GAGCAUGCCGUCAAUAAUA	1812
	CAAGAGAACUCAUCAUAAAG	1813
	CCAAAGUGCUGCAGUACUA	1814
	UAAGAAAGCUAUCCAGAUU	1815
FKBP1A	Sense	
	GAAACAAGCCCUUUAAGUU	1816
	GAAUUACUCUCCAAGUUGA	1817
	CAGCACAAGUGGUAGGUUA	1818
	GUUGAGGACUGAAUUACUC	1819
	GAUGGCAGCUGUUUAAAUG	1820
	GAGUAUCCUUUCAGUGUUA	1821
TNFRSF1A	Sense	
	CAAAGGAACCUACUUGUAC	1822
	GGAACCUACUUGUACAAUG	1823
	GAACCUACUUGUACAAUGA	1824
	GAGUGUGUCUCCUGUAGUA	1825

IL1R1	Sense	
	GGACAAGAAUCAUUGGAUA	1826
	GAACAAGCCUCCAGGAUUC	1827
	GGACUUGUGUGCCCUUAUA	1828
	GAACACAAAGGCACUAUAA	1829
IRAK1	Sense	
	CGAAGAAAGUGAUGAAUUU	1830
	GCUCUUUGCCCAUCUCUUU	1831
	UGAAAGACCUGGUGGAAGA	1832
	GCAAUUCAGUUUCUACAUC	1833
TRAF2	Sense	
	GAAGACAGAGUUAUUAAC	1834
	UCACGAAGACAGAGUUAUU	1835
	AGACAGAGUUAUUAACCA	1836
	CACGAAGACAGAGUUAUUA	1837
	GCUGAAGCCUGUCUGAUGU	1838
TRAF6	Sense	
	CAAUGAUCUGAGGCAGUU	1839
	GUUCAUAGUUUGAGCGUUA	1840
	GGAGAAACCUGUUGUGAUU	1841
	GGACAAAGUUGCUGAAAUC	1842
	CAAUGAUCUGAGGCAGUU	1843
	GGAGAAACCUGUUGUGAUU	1844
	GGACAAAGUUGCUGAAAUC	1845
	GUUCAUAGUUUGAGCGUUA	1846
TRADD	Sense	
	UGAAGCACCUUGAUCUUUG	1847
	GGGCAGCGCAUACCGUUU	1848
	GAGGAGCGCUGUUUGAGUU	1849
	GGACGAGGAGCGCUGUUUG	1850
	GAGGAGCGCUGUUUGAGUU	1851
	GGAUGUCUCUCUCCUCUUU	1852
	GCUCACUCCUUUCUACUAA	1853
	UGAAGCACCUUGAUCUUUG	1854
FADD	Sense	
	GCACAGAUUUUCCAUUUC	1855
	GCAGUCCUCUUAUCCUAA	1856
	GAACUCAAGCUGCGUUUAU	1857
	GGACGAAUUGAGAUAAUUA	1858
IKBKE	Sense	
	UAAGAACACUGCUCAUGAA	1859
	GAGGCAUCCUGAAGCAUUA	1860
	GAAGGCGGCUGCAGAACUG	1861

	GGAACAAGGAGAUCAUGUA	1862
IKBKG	Sense	
	CUAUCGAGGUCGUUAAAUU	1863
	GAAUGCAGCUGGAAGAUCU	1864
	GCGGCGAGCUGGACUGUUU	1865
	CCAGACCGAUGUGUAUUUA	1866
TNFRSF5	Sense	
	GGUCUCACCUCGCUAUGGU	1867
	GAAAGCGAAUCCUAGACA	1868
	GCACAAACAAGACUGAUGU	1869
	GAAGGGCACCUCAGAAACA	1870
	UCUCCCAACUUGUAUUAAA	1871
RELA	Sense	
	UCAAGUGUCUCCAUCUAG	1872
	UCAAGUGCCUUAUAGUAG	1873
	GGAGUACCCUGAGGCUAUA	1874
	GAUGAGAUCUCCUACUGU	1875
ARHA	Sense	
	GAGCUGGGCUAAGUAAAUA	1876
	GACCAAAGAUGGAGUGAGA	1877
	GGAAGAAACUGGUGAUUGU	1878
	GGCUGUAACUACUUUAUAA	1879
CDC42	Sense	
	GGACAUUUGUUUGCCAUUU	1880
	GGAGAACCAUAUACUCUUG	1881
	GAACCAAUGCUUUCUCAUG	1882
	GAAGACCUGUUAUGUAGAG	1883
	GAUCAAGAAUUGCAAUAUC	1884
	GAAAAGGGGUGACCUAGUA	1885
	UGACAAACCUUAUGGAAAA	1886
ROCK1	Sense	
	GGAAUGAGCUUCAGAUGCA	1887
	GGACACAGCUGUAAGAUUG	1888
	GACAAGAGAUUACAGAUAA	1889
	GAAGAAACAUUCCCUAUUC	1890
PAK1	Sense	
	GAGGGUGGUUUAUGAUUAA	1891
	CAACAAAGAACAUCACUA	1892
	GAAGAAAUAUACACGGUUU	1893
	UACAUGAGCUUUACAGAUAA	1894
PAK2	Sense	

	GGUAGGAGAUGAAUUGUUU	1895
	AGAAGGAACUGAUCAUUAA	1896
	CUACAGACCUCCAAUAUCA	1897
	GAAACUGGCCAAACCGUUA	1898
PAK3	Sense	
	GAUUAUCGCUGCAAAGGAA	1899
	GAGAGUGCCUGCAAGCUUU	1900
	GACAAGAGGUGGCCAUAAA	1901
	UUAAAUCGCUGUCUUGAGA	1902
PAK4	Sense	
	ACUAAGAGGUGAACAUGUA	1903
	GAUCAUGAAUGUCCGAAGA	1904
	GAUGAGACCCUACUACUGA	1905
	CAGCAAAGGUGCCAAAGAU	1906
PAK6	Sense	
	UAAAGGCAGUUGUCCACUA	1907
	GAAGGGACCUGCUUUCUUG	1908
	GCAAAGACGUCCCUAAGAG	1909
	CCAAUGGGCUGGCUGCAAA	1910
PAK7	Sense	
	GAGCACGGCUUUAUAAGU	1911
	CAAACUCCGUUAUGAUUA	1912
	GGAUAAAGUUGUCUGAUUU	1913
	GGAAAUGCCUCCAUAUAUA	1914
HDAC1	Sense	
	GGACAUCGCUGUGAAUUGG	1915
	AGAAAGAAGUCACCGAAGA	1916
	GGACAAGGCCACCCAAUGA	1917
	CCACAGCGAUGACUACAUU	1918
HDAC2	Sense	
	GCUGUAAAAUUAUGGCUUA	1919
	GCAAAGAAAGCUAGAAUUG	1920
	CAUCAGAGAGUCUUAUAUA	1921
	CCAAUGAGUUGCCAUUAUA	1922
CREBBP	Sense	
	GGCCAUAGCUUAAUUAUUC	1923
	GCACAGCCGUUUACCAUGA	1924
	GGACAGCCCUUUAGUCAAG	1925
	GAACUGAUUCCUGAAAUA	1926
BTRC	Sense	
	CACAUAAACUCGUAUCUUA	1927

	GAGAAGGCACUCAAGUUUA	1928
	AGACAUAGUUUACAGAGAA	1929
	GCAGAGAGAUUUUCAUAACU	1930
RIPK2	Sense	
	GAACAUACCUGUAAAUCAU	1931
	GGACAUCGACCUGUUUAUA	1932
	UAAAUAGAACUCCUACAUAG	1933
	GGAAUUUAUCUCUGAACAU	1934
VAV1	Sense	
	GCAGAAAUACAUCUACUAA	1935
	GCUAUGAGCUGUUCUCAA	1936
	CGACAAAGCUCUACUCAUC	1937
	GCUCAACCCUGGAGACAUU	1938
VAV2	Sense	
	GGACAAGACUCGCAGAUUU	1939
	GCUGAGCGCUUUGCAAUAA	1940
	CAAGAAGUCUCACGGGAAA	1941
	UCACAGAGGCCAAGAAAUU	1942
GRB2	Sense	
	UGGAAGCCAUCGCCAAAUA	[[1942]] 432
	CAUCAGUGCAUGACGUUUA	1943
	UGAAUGAGCUGGUGGAUUA	1944
	UGCCAAAACUUACCUAUAA	1945
PLCG1	Sense	
	GAGCUGCACUCCAAUGAGA	1946
	GAAACCAAGCCAUAUAUGA	1947
	CCAAGGAGCUACUGACAUU	1948
	AGAGAAACAUGGCCCAAUA	1949
ITGB1	Sense	
	CCACAGACAUUUACAUUAA	1950
	GAAGGGAGUUUGCUAAAUA	1951
	GAACAGAUCUGAUGAAUGA	1952
	CAAGAGAGCUGAAGACUAU	1953
ITGA4	Sense	
	GCAUAUAUAUUCAGCAUUG	1954
	CAACUUGACUGCAGUAUUG	1955
	GAACUUAACUUUCCAUGUU	1956
	GACAAGACCUGUAGUAAUU	1957
STAT1	Sense	
	AGAAAGAGCUUGACAGUAA	1958
	GGAAGUAGUUCACAAAUA	1959

	UGAAGUAUCUGUAUCCAAA	1960
	GAGCUUCACUCCCUUAGUU	1961
KRAS2	Sense	
	UAAGGACUCUGAAGAUGUA	1962
	GACAAAGUGUGUAAUUAUG	1963
	GCUCAGGACUUAGCAAGAA	1964
	GAAACUGAAUACCUAAGAU	1965
	GAAACUGAAUACCUAAGAU	1966
	UAAGGACUCUGAAGAUGUA	1967
	GACAAAGUGUGUAAUUAUG	1968
	GCUCAGGACUUAGCAAGAA	1969
HRAS	Sense	
	CCAUCCAGCUGAUCCAGAA	1970
	GAACCCUCCUGAUGAGAGU	1971
	GAGGACAUCCACCAGUACA	1972
BRAF	Sense	
	GAUUAGAGACCAAGGAUUU	2410
	CCACUGAUGUGUGUUAUU	1973
	CAUAGAACCUGUCAAUU	1974
	GAAGACAGGAAUCGAAUGA	1975
ELK1	Sense	
	GAUGUGAGUAGAAGAGUUA	[[1975]] 2411
	GGAAGAAUUUGUACCAUUU	1976
	GAACGACCUUUCUUUCUUU	1977
	GGAGUCAUCUCUCCUAUA	1978
RALGDS	Sense	
	GGAGAAGCCUCACCUCUUG	1979
	GCAGAAAGGACUCAAGAUU	1980
	GAGAACAACUACUCAUUGA	1981
	GAACUUCUCGUCACUGUAU	1982
PRKCA	Sense	
	GGAUUGUUCUUUCUUAUA	1983
	GAAGGGUUCUCGUAUGUCA	1984
	GAAGAAGGAUGUGGUGAUU	1985
	GGACUGGGAUCGAACAACA	1986
MAP2K4	Sense	
	GGACAGAAGUGGAAAUUU	1987
	UCAAGAGGUGAACAUAUA	1988
	GACCAAUUCUCAGUUGUUU	1989
	GGAGAAUGGUGCUGUUUAU	1990
MAP2K7	Sense	

	GAAGAGACCAAAGUAUAAU	1991
	GAAGACCGGCCACGUCAUU	1992
	GGAAGAGACCAAAGUAUAA	1993
	GCAUUGAGAUUGACCAGAA	1994
	UGAGAGAACGAGAAAAGUUG	1995
	GUGAAACCCUGUCUGCAUU	1996
	GGAUCUCUCUCAACAACUA	1997
	ACAACUAGGUGAACACAUA	1998
MAPK8	Sense	
	UCACAGUCCUGAAACGAUA	1999
	GAUUGGAGAUUCUACAUUC	2000
	GCUCAUGGAUGCAAAUCUU	2001
	GAAGCUAAGCCGACCAUUU	2002
MAPK9	Sense	
	AAAGAGAGCUUAUCGUGAA	2003
	GAUGAUAGGUUAGAAAUAG	2004
	ACAAAGAAGUCAUGGAUUG	2005
	GGAGCUGGAUCAUGAAAGA	2006
AIF1	Sense	
	GAAAAGGGAUGAUGGGAUU	2007
	CCUAGACGAUCCCAAUAU	2008
	GAGCCAAACCAGGGAUUUA	2009
	UGAAACGAAUGCUGGAGAA	2010
	UCACUCACCCAGAGAAUA	2011
	CCAAGAAAGCUAUCUCUGA	2012
	AGACUCACCUAGAGCUAAA	2013
BBC3	Sense	
	CCUGGAGGGUCCUGUACAA	2014
	GAGCAAAUGAGCCAAACGU	2015
	GGAGGGUCCUGUACAAUCU	2016
	GACUUUCUCUGCACCAUGU	2017
BCL2L1	Sense	
	CCAGGGAGCUUGAAAGUUU	2018
	AAAGUGCAGUUCAGUAAUA	2019
	GAGAAUCACUAACCAGAGA	2020
	GAGCCCAUCCCUAUUAUAA	2021
BCL2L11	Sense	
	GAGACGAGUUUAACGCUUA	2022
	AAAGCAACCUUCUGAUGUA	2023
	CCGAGAAGGUAGACAAUUG	2024
	GCAAAGCAACCUUCUGAUG	2025
	AGACAGAGCCACAAGGUAA	2026
	GCAAGGAGGUUAGAGAAAU	2027

BID	CAAGGAGGUUAGAGAAUA	2028
	UCUUACGACUGUUACGUUA	2029
	Sense	
	GAAGACAUCAUCCGGAUA	2030
	CAACAGCGUUCUAGAGAA	2031
BIRC2	GAAUUGGGAUGGACUGAAC	2032
	ACGAUGAGCUGCAGACUGA	2033
	Sense	
	GAAAGAAGCCUGCAUAUAA	2034
	GAAAUUGACUCUACAUUGU	2035
BIRC3	ACAAAUAGCACUUAGGUUA	2036
	GAAUACACCUGUGGUUAAA	2037
	Sense	
	GGAGAUGCCUGCCAUAUAAA	2038
	UCAUGAUCUUGUGUUAGA	2039
BIRC4	GAAAGAACAUGUAAAGUGU	2040
	GAAGAAAGAACAUGUAAAG	2041
	Sense	
	GUAGAUAGAUGGCAUAUUG	2042
	GAGGAGGGCUAACUGAUUG	2043
BIRC5	GAGGAACCCUGCCAUGUAU	2044
	GCACGGAUCUUUACUUUUG	2045
	Sense	
	GGCGUAAGAUGAUGGAUUU	2046
	GCAAAGGAAACCAACAAUA	2047
BRCA1	GCACAAAGCCAUUCUAAGU	2048
	CAAAGGAAACCAACAAUA	2049
	Sense	
	CCAUACAGCUUCAUAAAUA	2050
	GAAGAGAACUUAUCUAGUG	2051
CARD4	GAAGUGGGCUCCAGUAUUA	2052
	GCAAGAUGCUGAUUCAUUA	2053
	CCAUACAGCUUCAUAAAUA	2054
	Sense	
	GAAAGUUAUGUCAAGGAA	2055
CASP10	GAGCAACACUGGCAUAACA	2056
	UACAGAGAUUUGCCUAAA	2057
	GCGAAGAGCUGACCAAUA	2058
	Sense	
	CAAAGGGUUCUCUGUUUA	2059
	GAAUAGACCUCCUAAGUU	2060

	GAAGGCAGCUGGUAAUUAUC	2061
	GACAUGAUCUUCUUCUGA	2062
	GCACUCUUCUGUUCUUUA	2063
CASP2	Sense	
	GUUUUAAACUCUCCUUUGA	2064
	GCAAGGAGAUGUCUGAAUA	2065
	CAACUUCUCCUGAUCUUUAA	2066
	GCUCAAAGAUGUAAUGUAG	2067
CDKN1A	Sense	
	GAACAAGGAGUCAGACAUU	2068
	AAACUAGGCGGUUGAAUGA	2069
	GAUGGAACUUCGACUUUGU	2070
CFLAR	GUAACAGAUGGCACUUUG	2071
	Sense	
	GAUGUGUCCUCAUUAAUUU	2072
	GAAGAGAGAUACAAGAUGA	2073
	GAGCAUACCUGAAGAGAGA	2074
CLK2	GCUAUGAAGUCCAGAAAUU	2075
	Sense	
	GUGAAUAUGUGAAAUAGUG	2076
	AAAGCAUGCUAGAGUAUGA	2077
	UUAAGAAUGUGGAGAAGUA	2078
CLSPN	GAUAACAAGCUGACACAUU	2079
	Sense	
	GGACGUAAUUGAUGAAGUA	2080
	GCAGAUGGGUUCUUAAAUG	2081
	CAAAUGAGGUUGAGGAAAU	2082
CSNK2A1	GGAAAUACCUGGAGGAUGA	2083
	Sense	
	GAUCCACGUUUCUUAUGUA	2084
	GCAUUUAGGUGGAGACUUC	2085
	GAUGUACGAUUUAUAGUUUG	2086
CTNNB1	UGAAUUAGAUCACGUUUC	2087
	Sense	
	GCACAAGAAUGGAUCACAA	2088
	GCUGAAACAUGCAGUUGUA	2089
	GUACGUACCAUGCAGAAUA	2090
CXCR4	GAACUUGCAUUGUGAUUGG	2091
	Sense	
	GAAGCAUGACGGACAAGUA	2092
	GAACAUUCCAGAGCGUGUA	2093

	GUUCUUAGUUGCUGUAUGU	2094
	CAUCAUGGUUGGCCUUAUC	2095
CXCR6	Sense	
	GGAACAAACUGGCAAAGCA	2096
	GAUCAGAGCAGCAGUGAAA	2097
	GGGCAAAACUGAAUUAUAA	2098
	GAUCUCAGGUUCUCCUUGA	2099
DAXX	Sense	
	CUACAGAUCUCCAAUGAAA	2100
	GCUACAAGCUGGAGAAUGA	2101
	GGAAACAGCUAUGUGGAAA	2102
	GGAGUUGGAUCUCUCAGAA	2103
GAS41	Sense	
	GUAGUAAGCUAAACUGAAA	2104
	GACAAUAUGUUCAAGAGAA	2105
	GACAACAUCUCGUCAGCUA	2106
	UAUAUGAUGUGUCCAGUAA	2107
GTSE1	Sense	
	CAAAGAAGCUCACUUAACUG	2108
	GAACAGCCCUAAAGUGGUU	2109
	GAACAUGGAUGACCCUAAG	2110
	GGGCAAAGCUAAAUCAAGU	2111
HDAC3	Sense	
	GGAAAGCGAUGUGGAGAUU	2112
	CCAAGACCGUGGCCUUAUU	2113
	AAAGCGAUGUGGAGAUUUA	2114
	GUGAGGAGCUUCCCUAUAG	2115
HDAC5	Sense	
	GAAUUCUCUUGUCGAAGU	2116
	GUUAUUAGCACCUUUAAGA	2117
	GGAGGGAGGCCAUGACUUG	2118
	CAGGAGAGCUCAAGAAUGG	2119
	GAUAUGGAUUUCAGUUA	2120
	GGAAGUCGGUGCCUUGGUU	2121
	GGAAGGAGAGGACUGGUUU	2122
HEC	Sense	
	GCAGAUACUUGCACGGUUU	2123
	GAGUAGAACUAGAAUGUGA	2124
	GCGAAUAAAUCAUGAAAGA	2125
	GAAGAUGGAAUUAUGCAUA	2126
HIST1H2	Sense	

AA	GGCAAUGCGUCUCGCGAUA	2127
	GAUCCGCAAUGAUGAGGAA	2128
	GCAAUGCGUCUCGCGAUAA	2129
	GAGGAACUCAUAAGCUUU	2130
LMNB1	Sense	
	AAUAGAAGCUGUGCAAUUA	2131
	CAACUGACCUCaucUGGAA	2132
	GAAGGAAUCUGAUCUUAU	2133
	GGGAAGGGUUUCUCAUUA	2134
LMNB2	Sense	
	GGAGGUUCAUUGAGAAUUG	2134
	GGCAAUAGCUCACCGUUUA	2135
	CAAUACGCUUAGCUGUGU	2136
	GGAGAU CGCCUACAAGUUC	2137
MYB	Sense	
	GCAGAAACACUCCAAUUUA	2138
	GUAAAUACGUGAAUGCAUU	2139
	GCACUGAACUUUUGAGAU	2140
	GAAGAACAGUCAUUUGAUG	2141
MYT1	Sense	
	GAGGUGAGCUGUUAAAUCA	2142
	GCAGGGUGAUUCCUAAUA	2143
	GGGAGAAGAUUUUAAUUG	2144
	CAACUUCUCUCCUGAACUU	2145
NFKBIB	Sense	
	GGACACGGCACUGCACUUG	2146
	GCACUUGGCUGUGAUUCAU	2148
	GAGACGAGGGCGAUGAAUA	2149
	CAUGAACCCUCCUGGAUU	2150
NFKBIA	Sense	
	GAACAUGGACUUGUAUUAU	2151
	GAUGUGGGGUGAAAAGUUA	2152
	GGACGAGAAAGAUCAUUGA	2153
	AGGACGAGCUGCCCUAUGA	2154
NFKBIE	Sense	
	GAAGGGAAGUUUCAGUAAC	2155
	GGAAGGGAAGUUUCAGUAA	2156
	GGAAACUGCUGCUGUGUAC	2157
	GAACCAACCACUCAUGGAA	2158
NUMA1	Sense	
	GGGAACAGUUUGAAUAUAA	2159

	GCAGUAGCCUGAAGCAGAA	2160
	CGAGAAGGAUGCACAGUA	2161
	GCAAGAGGCUGAGAGGAAA	2162
NUP153	Sense	
	GAAGACAAUGAAAGCUAA	2163
	GAUAAAGACUGCUGUUAGA	2164
	GAGGAGAGCUCUAAUAUUA	2165
	GAGGAAGCCUGAUUAAAGA	2166
OPA1	Sense	
	GAAAGAGCAUGAUGACAU	2167
	GAGGAGAGCUCUAAUUAUGU	2168
	GAAACUGAAUGGAAGAAUA	2169
	AAAGAAGGCUGUACCGUUA	2170
PARVA	Sense	
	CUACAUGUCUUUGCUCUUA	2171
	GCUAAGUCCUGUAAGAAUA	2172
	CAAAGGCAAUGUACUGUUU	2173
	GAACAAUGGUGGAUCCAAA	2174
PIK3CG	Sense	
	AAGUUCAGCUUCUCUAUUA	2175
	GAAGAAAUCUCUGAUGGAU	2176
	GAACACCUUUACUCUAUAA	2177
	GCAUGGAGCUGGAGAACUA	2178
PRKDC	Sense	
	GAUGAAAGCUCUAAAGAUG	2179
	GAAAGGAGGUUCUAAACUA	2180
	GGAAGAAGCUCAUUUGAUU	2181
	GCAAAGAGGUGGCAGUUA	2182
RASA1	Sense	
	GGAAGAAGAUCCACAUGAA	2183
	GAACAUACUUUCAGAGCUU	2184
	GAACAAUCUUUGCUGUAUA	2185
	UAACAGAACUGCUUCAACA	2186
SLC9A1	Sense	
	GAAGAGAUCCACACACAGU	2187
	UCAUGAGCUGCUGCACAU	2188
	GAAGAUAGGUUCCAUGUG	2189
	GAAUUACCCUCCUCAUCU	2190
TEGT	Sense	
	CUACAGAGCUUCAGUGUGA	2191
	GAACAUUUUGAUCGAAAG	2192

TERT	GAGCAAACCUAGAUAAAGGA	2193
	GCAUUGAUCUCUUCUUAGA	2194
	Sense	
	GGAAGACAGUGGUGAACUU	2195
	GCAAAGCAUUGGAAUCAGA	2196
	GAGCUGACGUGGAAGAUGA	2197
	GAACGGGCCUGGAACCAUA	2198
TNFRSF6	Sense	
	GAUACUAACUGCUCUCAGA	2199
	GAAAGAAUGGUGUCAAUGA	2200
	UCAUAAUGUCCCAUGUAA	2201
	UCAUGAAUCUCCAACCUUA	2202
	GAUGUUGACUUGAGUAAAU	2203
TOP1	Sense	
	GAAAGGAAAUGACUAAUGA	2204
	GAAGAAGGCUGUUCAGAGA	2205
	GGAAGUAGCUACGUUCUUU	2206
	GGACAUAAAGUGGAAAGAAG	2207
TOP2A	Sense	
	GAAAGAGUCCAUCAGAUUU	2208
	CAAACUACAUUGGCAUUUA	2209
	AAACAGACAUGGAUGGAUA	2210
	CGAAAGGAAUGGUUAACUA	2211
TOP3A	Sense	
	CCAGAAAUCUCCACAGAA	2212
	GAAACUAUCUGGAUGUGUA	2213
	CCACAAAGAUGGUAUCGUA	2214
	GGAAAUGGCUGUGGUAACA	2215
TOP3B	Sense	
	GAGACAAGAUGAAGACUGU	2216
	GCACAUGGGCUGCGUCUUU	2217
	CCAGUGCGCUUCAAGAUGA	2218
	GAACAUCUGCUUUGAGGUU	2219
WEE1	Sense	
	GGUAUUGCCUUGUGAAUUU	2220
	GCAGAACAAUUACGAAUAG	2221
	GUACAUAGCUGUUUGAAAU	2222
	GCUGUAAACUUGUAGCAUU	2223

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 40 of 42

Please replace the paragraph on page 146, spanning lines 8-12, and replace it with the following amended paragraph:

the interleukin 4 receptor gene

(SEQ. ID NO. 2224: UAGAGGUGCUCAUUC AUUU,
SEQ. ID NO. 2225: GGUAUAAGCCUUUCCAAGA,
SEQ. ID NO. ~~[[2225]]~~ 2412: ACACACAGCUGGAAGAAAU,
SEQ. ID NO. 2226: UAACAGAGCUUCCUUAGGU),

Please replace the paragraph on page 148, spanning lines 8-19, and replace it with the following amended paragraph:

activin A receptor IB

(SEQ. ID NO. 2255: ACAAGACGCUCCAGGAUCU,
SEQ. ID NO. ~~[[2254]]~~ 2413: GCAACAGGAUCGACUUGAG,
SEQ. ID NO. ~~[[2255]]~~ 2414: GAAGCUGCGUCCCAACAUC,
SEQ. ID NO. 2256: GCAUAGGCCUGUAAUCGUA,
SEQ. ID NO. 2257: UCAGAGAGUUCGAGACAAA,
SEQ. ID NO. 2258: UGCGAAAGGUUGUAUGUGA,
SEQ. ID NO. 2259: GCAACAGGAUCGACUUGAG,
SEQ. ID NO. 2260: GAAUAGCGUUGUGUGUUUAU,
SEQ. ID NO. 2261: UGAAUAGCGUUGUGUGUUA,
SEQ. ID NO. 2262: GGGAUCAGUUUGUUGAAUA,
SEQ. ID NO. 2263: GAGCCUGAAUCAUCGUUUA,),

Please replace the paragraph on page 150, spanning lines 22-26, and replace it with the following amended paragraph:

NOS3

(SEQ. ID NO. 2308: UGAAGCACCUUGGAGAAUGA,

Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 41 of 42

SEQ. ID NO. 2309: CGGAACAGCACAAGAGUUA,
SEQ. ID NO. 2310: GGAAGAAGACCUUUAAGA,
SEQ. ID NO. [[2309]] 2415: GCACAAGAGUUAUAAGAUC),

Please replace the paragraph on page 150, spanning lines 28-32, and replace it with the following amended paragraph:

ARH

(SEQ. ID NO. [[2310]] 2416: CGAUACAGCUUGGCACUUU,
SEQ. ID NO. 2311: GAGAAGCGCUGCCCUGUGA,
SEQ. ID NO. 2312: GAAUCAUGCUGUUCUCUUU,
SEQ. ID NO. 2313: GGAGUAACCGGACACCUUA),

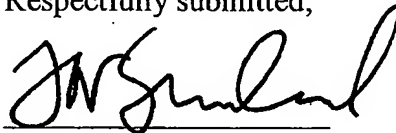
Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
May 20, 2004
Page 42 of 42

REMARKS

Applicants have amended the specification by assigning SEQ. ID NOS. to sequences lacking them, and changing sequence ID numbers to correct instances where more than one sequence was assigned to a single SEQ. ID NO. Applicants have also amended the specification to correct four typographical errors consisting of dual "T" residues at the end of each of SEQ. ID NOS. 1413-1416. The amendments add no new matter. Applicants respectfully request entry of the amendments. Substitute specification pages reflecting the amendments would be provided to the Examiner upon the Examiner's request.

No fee is required in connection with the filing of this Preliminary Amendment. If any fee is deemed necessary, authorization is hereby given to charge the amount of any such fee to Deposit Account No. 11-0171.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Tor Smeland", written over a horizontal line.

Tor Smeland
Registration No.: 43,131
Attorney for Applicants

Kalow & Springut LLP
Telephone No.: (212) 813-1600

ITEM 3

IN THE UNITED STATES PATENT AND TRADEMARK OFFICEApplicant: KHVOROVA *et al.* Examiner: To be assigned

Serial No.: 10/714,333 Group Art Unit:

Filed: November 14, 2003

For: Functional and Hyperfunctional siRNA

Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

June 30, 2004

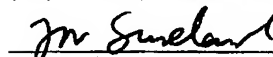
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450**AMENDMENT DIRECTING ENTRY OF SEQUENCE LISTING**

Sir:

This amendment directs entry into this application of the sequence listing filed by
Express Mail on June 29, 2004.

Certificate of Transmission Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being facsimile transmitted to the United States Patent and Trademark Office (Fax. No. (703) 746-4060) on the date shown below.


(Signature)TOR SMELAND
(Printed Name of Person Signing Certificate)30 JUNE 2004
(Date)

Applicant: KHVOROVA et al.
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
June 30, 2004
Page 2 of 4

Amendments to the Specification:

Please insert the sequence listing filed on June 29, 2004, after page 159 and before page 160.

REMARKS

Applicants herein request entry of the sequence listing filed on June 29, 2004 into the instant application. In response to a Notice to Comply with Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Sequence Disclosures (Filing Date Granted) mailed February 12, 2004, Applicants filed a sequence listing on CD-ROM, in lieu of a paper sequence listing, on June 29, 2004 by Post Office to Addressee Express Mail, addressed to Box Missing Parts, with a petition and fee for a three-month extension of time. Applicants herein request entry of the sequence listing filed, which listing is a text file entitled "13499US.txt" on the disk entitled "FILE COPY – SEQUENCE LISTING COPY 1 OF 3 (PAPER COPY)".

Support for the amendment can be found in the application as filed, for example, in Table_12.txt, Table_13.txt, Table_14.txt, and Table_15.txt, which are tables related to nucleotide sequences that were filed on CD-ROM on the application's filing date, in duplicate, on a compact disk labeled "DOCKET 13499, PATENT APPLICATION, DISK 1 OF 1, COPY 1 OF 2." The amendment adds no new matter. Applicants respectfully request entry of the amendment.

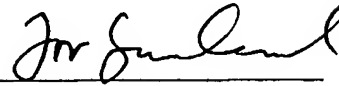
Conclusion

No fee is believed to be due with respect to the filing of this amendment. If any additional fees are due, or an overpayment has been made, please charge, or credit, our Deposit Account No. 11-0171 for such sum.

If the Examiner has any questions regarding the present application, the Examiner is cordially invited to contact Applicant's attorney at the telephone number provided below.

Applicant: KHVOROVA et al.
Serial No.: 10/714,333
Filing Date: November 14, 2003
Preliminary Amendment
June 30, 2004
Page 4 of 4

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Tor Smeland", written over a horizontal line.

Tor Smeland
Registration No.: 43,131
Attorney for Applicant

Kalow & Springut LLP
Telephone No.: (212) 813-1600

ITEM 4

PATENT

DOCKET 13499USC2

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: KHVOROVA *et al.* Examiner: To be assigned
Serial No.: 10/714,333 Group Art Unit: 1646
Filed: November 14, 2003
For: Functional and Hyperfunctional siRNA
Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

April 21, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

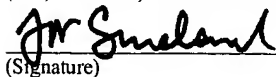
SUPPLEMENTAL PRELIMINARY AMENDMENT

Dear Sir:

Prior to examination on the merits, please enter the following amendments in the above-identified application.

Certificate of Transmission Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being facsimile transmitted to the United States Patent and Trademark Office (Fax. No. (703) 872-9306) on the date shown below.


(Signature)

TOR SMEALAND
(Printed Name of Person Signing Certificate)

21 APRIL 2005
(Date)

Amendments to the Claims:

This listing will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (original) A method for selecting siRNA comprising selecting an siRNA molecule of 19 – 25 nucleoside bases, said method comprising:
 - (a) selecting a target gene;
 - (b) measuring the functionality of sequences of 19 – 25 nucleotides in length that are substantially complementary to a stretch of nucleotides of the target sequence, wherein said functionality is dependent upon non-target specific criteria.
2. (original) The method according to claim 1 wherein said functionality is determined by applying one of the following formulas:

Formula I = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$

Formula II = $-(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$

Formula III = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3;$

Formula IV = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$

Formula V = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$

Formula VI = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$

$$\text{Formula VII} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 \\ + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2;$$

wherein in Formulas I – VII:

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at positions 15 – 19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$

wherein

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;

$C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

$GC_{15-19} =$ the number of G and C bases within positions 15 – 19 of the sense strand or within positions 15 –18 if the sense strand is only 18 base pairs in length;

GC_{total} = the number of G and C bases in the sense strand;

T_m = 100 if the targeting site contains an inverted repeat longer than 4 base pairs,
otherwise its value is 0; and

X = the number of times that the same nucleotide repeats four or more times in a row.

3. (original) A method of gene-silencing comprising selecting an siRNA according to claim 2 and introducing it into a cell.
4. (original) The method according to claim 3 wherein said introducing is by allowing passive uptake of the siRNA.
5. (original) The method according to claim 3, wherein said introducing is through the use of a vector.
6. (original) A method for developing an siRNA algorithm for selecting siRNA, said method comprising:
 - (a) selecting a set of siRNA;
 - (b) measuring the gene silencing ability of each siRNA from said set;
 - (c) determining the relative functionality of each siRNA;
 - (d) determining the amount of improved functionality by the presence or absence of at least one variable selected from the group consisting of the total GC content, melting temperature of the siRNA, GC content at positions 15 –19, the presence or absence of a particular nucleotide at a particular position and the number of times that the same nucleotide repeats within a given sequence; and
 - (e) developing an algorithm using the information of step (d).
7. (original) A method of selecting an siRNA with improved functionality, said method comprising using the algorithm of claim 6.

8. (original) A method of selecting hyperfunctional siRNA, said method comprising using at least one functional siRNA, wherein at least one said functional siRNA has been selected according to the method of claim 7 and measuring the silencing ability of said at least one functional siRNA, wherein silencing ability is measured at a concentration of less than 1 nanomolar siRNA.

9.-18. (canceled)

19. (original) A kit, wherein said kit is comprised of at least two siRNA, wherein said at least two siRNA comprise a first optimized siRNA and a second optimized siRNA, wherein said first optimized siRNA and said second optimized siRNA are optimized according to one of the following formulas:

$$\text{Formula I} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) \\ + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula II} = -(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$$

$$\text{Formula III} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3;$$

$$\text{Formula IV} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) \\ + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula V} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula VI} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$$

$$\text{Formula VII} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 + (A_3) * 3 \\ + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2;$$

wherein in Formulas I – VII:

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at positions 15 – 19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-$
 $9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4$
 $+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+$
 $18*A_3+19*U_{10}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*$
 $C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+$
 $(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-$
 $13.7)*G_{13}+(-25.9)*G_{19}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$

wherein

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;

$C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

$GC_{15-19} =$ the number of G and C bases within positions 15 – 19 of the sense strand or within positions 15 –18 if the sense strand is only 18 base pairs in length;
 $GC_{total} =$ the number of G and C bases in the sense strand;

Applicants: KHVOROVA *et al.*
Serial No.: To be assigned
Filing Date: Filed herewith
Supplemental Preliminary Amendment
April 21, 2005
Page 11 of 12

$T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs, otherwise its value is 0; and

$X =$ the number of times that the same nucleotide repeats four or more times in a row.

Applicants: KHVOROVA *et al.*
Serial No.: To be assigned
Filing Date: Filed herewith
Supplemental Preliminary Amendment
April 21, 2005
Page 12 of 12

REMARKS

Claims 1-19 are pending in the application. This Supplemental Preliminary Amendment cancels claims 9-18 without disclaimer. Applicants reserve the right to file one or more divisional or continuation applications based on the canceled claims. The amendment adds no new matter. Applicants respectfully request entry of this amendment.

No fee is required in connection with the filing of this Preliminary Amendment. If any fee is deemed necessary, or overpayment has been made, please charge, or credit, Deposit Account No. 11-0171.

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Tor Smeland", written over a horizontal line.

Tor Smeland
Registration No.: 43,131
Attorney for Applicants

Kalow & Springut LLP
Telephone No.: (212) 813-1600

ITEM 5

To:
DAVID KALOW
KALLOW & SPRINGUT LLP
488 MADISON AVE. 19TH FL.
NEW YORK, NY 10022

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

Applicant's or agent's file reference 13499 PCT	Date of Mailing (day/month/year) FOR FURTHER ACTION See paragraphs 1 and 4 below
International application No. PCT/US03/36787	International filing date (day/month/year) 14 November 2003 (14.11.2003)
Applicant DHARMA CON INC.	

1. ☒ The applicant is hereby notified that the international search report has been established and is transmitted herewith.

Filing of amendments. The applicant is entitled to file amendments under Article 17(2)(a) if so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

Where? Directly to the International Bureau of WIPO, 34, chemin des Colombettes
1211 Geneva 20, Switzerland, Facsimile No.: (41-22) 740.14.35

For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ With regard to the protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

- ☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.
☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. Reminders

Shortly after 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90 bis.1 and 90 bis.3, respectively, before the completion of the technical preparations for international publication.

Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months.

See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer <i>Janet L. Epps-Ford</i> Janet L. Epps-Ford, Ph.D. Telephone No. 571-272-0547
---	---

To:
DAVID KALOW
KALLOW & SPRINGUT LLP
488 MADISON AVE. 19TH FL.
NEW YORK, NY 10022

PCT

NOTIFICATION OF TRANSMITTAL OF THE INTERNATIONAL SEARCH REPORT OR THE DECLARATION

(PCT Rule 44.1)

Date of Mailing (day/month/year) 25 FEB 2005	
Applicant's or agent's file reference 13499 PCT	FOR FURTHER ACTION See paragraphs 1 and 4 below
International application No. PCT/US03/36787	International filing date (day/month/year) 14 November 2003 (14.11.2003)
Applicant DHARMACON INC.	

1. ☒ The applicant is hereby notified that the international search report has been established and is transmitted herewith.

Filing of amendments and statement under Article 19:

The applicant is entitled, if he so wishes, to amend the claims of the international application (see Rule 46):

When? The time limit for filing such amendments is normally two months from the date of transmittal of the international search report.

Where? Directly to the International Bureau of WIPO, 34, chemin des Colombettes
1211 Geneva 20, Switzerland, Facsimile No.: (41-22) 740.14.35

For more detailed instructions, see the notes on the accompanying sheet.

2. ☐ The applicant is hereby notified that no international search report will be established and that the declaration under Article 17(2)(a) to that effect is transmitted herewith.

3. ☐ With regard to the protest against payment of (an) additional fee(s) under Rule 40.2, the applicant is notified that:

☐ the protest together with the decision thereon has been transmitted to the International Bureau together with the applicant's request to forward the texts of both the protest and the decision thereon to the designated Offices.

☐ no decision has been made yet on the protest; the applicant will be notified as soon as a decision is made.

4. Reminders

Shortly after 18 months from the priority date, the international application will be published by the International Bureau. If the applicant wishes to avoid or postpone publication, a notice of withdrawal of the international application, or of the priority claim, must reach the International Bureau as provided in Rules 90 bis.1 and 90 bis.3, respectively, before the completion of the technical preparations for international publication.

Within 19 months from the priority date, but only in respect of some designated Offices, a demand for international preliminary examination must be filed if the applicant wishes to postpone the entry into the national phase until 30 months from the priority date (in some Offices even later); otherwise the applicant must, within 20 months from the priority date, perform the prescribed acts for entry into the national phase before those designated Offices.

In respect of other designated Offices, the time limit of 30 months (or later) will apply even if no demand is filed within 19 months.

See the Annex to Form PCT/IB/301 and, for details about the applicable time limits, Office by Office, see the *PCT Applicant's Guide*, Volume II, National Chapters and the WIPO Internet site.

Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer <i>Janet L. Epps-Ford</i> Janet L. Epps-Ford, Ph.D. Telephone No. 571-272-0547
---	---

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 13499 PCT	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US03/36787	International filing date (<i>day/month/year</i>) 14 November 2003 (14.11.2003)	(Earliest) Priority Date (<i>day/month/year</i>) 14 November 2002 (14.11.2002)
Applicant DHARMACON INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

☒ It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the Report

a. With regard to the **language**, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.

☐ the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

b. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international search was carried out on the basis of the sequence listing:

☐ contained in the international application in written form.

☒ filed together with the international application in computer readable form.

☐ furnished subsequently to this Authority in written form.

☐ furnished subsequently to this Authority in computer readable form.

☐ the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

☐ the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

2. ☐ **Certain claims were found unsearchable** (See Box I).

3. ☐ **Unity of invention is lacking** (See Box II).

4. With regard to the **title**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

☒ the text is approved as submitted by the applicant.

☐ the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the **drawings** to be published with the abstract is Figure No. 1

☒ as suggested by the applicant.

☐ because the applicant failed to suggest a figure.

☐ because this figure better characterizes the invention.

☐ None of the figures

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/36787

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/00

US CL : 435/6; 536/24.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 536/24.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002/0150945 A1 (FINNEY et al) 17 October 2002 (17.10.2002), see page 27-28.	1-19
A	KASIF et al. A computational framework for optimal masking in the synthesis of oligonucleotide microarrays. Nucleic Acids Research. 2002, Vol. 30, No. 20, full text version of article.	1-19
A	AMARZGUIOUI et al. Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes. Nuclei acids Research, 2000, Vol. 28, No. 21, pages 4113-4124, full text version.	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

19 January 2005 (19.01.2005)

Date of mailing of the international search report

25 FEB 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

Janet L. Epps-Ford, Ph.D.

Telephone No. 571-272-0547

Continuation of B. FIELDS SEARCHED Item 3:

CAplus, Medline, Biosis, USPatfull, Derwent, JPO, EPO

search terms: (SIRNA OR RNAI OR DSRNA) and (OPTIMIZATION OR OPTIMIZE OR OPTIMAL) and algorithm

NOTES TO FORM PCT/ISA/220

These Notes are intended to give the basic instructions concerning the filing of amendments under Article 19. The Notes are based on the requirements of the Patent Cooperation Treaty, the Regulations and the Administrative Instructions under that Treaty. In case of discrepancy between these Notes and those requirements, the latter are applicable. For more detailed information, see also the *PCT Applicant's Guide*, a publication of WIPO.

In these Notes, "Article," "Rule" and "Section" refer to the provisions of the PCT, the PCT Regulations and the PCT Administrative Instructions, respectively.

INSTRUCTIONS CONCERNING AMENDMENTS UNDER ARTICLE 19

The applicant has, after having received the international search report, one opportunity to amend the claims of the international application. It should however be emphasized that, since all parts of the international application (claims, description and drawings) may be amended during the international preliminary examination procedure, there is usually no need to file amendments of the claims under Article 19 except where, e.g. the applicant wants the latter to be published for the purposes of provisional protection or has another reason for amending the claims before international publication. Furthermore, it should be emphasized that provisional protection is available in some States only.

What parts of the international application may be amended ?

Under Article 19, only the claims may be amended.

During the international phase, the claims may also be amended (or further amended) under Article 34 before the International Preliminary Examining Authority. The description and drawings may only be amended under Article 34 before the International Preliminary Examining Authority.

Upon entry into the national phase, all parts of the international application may be amended under Article 28 or, where applicable, Article 41.

When ? Within 2 months from the date of transmittal of the international search report or 16 months from the priority date, whichever time limit expires later. It should be noted, however, that the amendments will be considered as having been received on time if they are received by the International Bureau after the expiration of the applicable time limit but before the completion of the technical preparations for international publication (Rule 46.1).

Where not to file the amendments ?

The amendments may only be filed with the International Bureau and not with the receiving Office or the International Searching Authority (Rule 46.2).

Where a demand for international preliminary examination has been/is filed, see below.

How ? Either by cancelling one or more entire claims, by adding one or more new claims or by amending the text of one or more of the claims as filed.

A replacement sheet must be submitted for each sheet of the claims which, on account of an amendment or amendments, differs from the sheet originally filed.

All the claims appearing on a replacement sheet must be numbered in Arabic numerals. Where a claim is cancelled, no renumbering of the other claims is required. In all cases where claims are renumbered, they must be renumbered consecutively (Administrative Instructions, Section 205(b)).

The amendments must be made in the language in which the international application is to be published.

What documents must/may accompany the amendments ?

Letter (Section 205(b)):

The amendments must be submitted with a letter.

The letter will not be published with the international application and the amended claims. It should not be confused with the "Statement under Article 19(1)" (see below, under "Statement under Article 19(1)").

The letter must be in English or French, at the choice of the applicant. However, if the language of the international application is English, the letter must be in English; if the language of the international application is French, the letter must be in French.

NOTES TO FORM PCT/ISA/220 (continued)

The letter must indicate the differences between the claims as filed and the claims as amended. It must, in particular, indicate, in connection with each claim appearing in the international application (it being understood that identical indications concerning several claims may be grouped), whether

- (i) the claim is unchanged;
- (ii) the claim is cancelled;
- (iii) the claim is new;
- (iv) the claim replaces one or more claims as filed;
- (v) the claim is the result of the division of a claim as filed.

The following examples illustrate the manner in which amendments must be explained in the accompanying letter:

1. [Where originally there were 48 claims and after amendment of some claims there are 51]:
"Claims 1 to 29, 31, 32, 34, 35, 37 to 48 replaced by amended claims bearing the same numbers; claims 30, 33 and 36 unchanged; new claims 49 to 51 added."
2. [Where originally there were 15 claims and after amendment of all claims there are 11]:
"Claims 1 to 15 replaced by amended claims 1 to 11."
3. [Where originally there were 14 claims and the amendments consist in cancelling some claims and in adding new claims]:
"Claims 1 to 6 and 14 unchanged; claims 7 to 13 cancelled; new claims 15, 16 and 17 added." or
"Claims 7 to 13 cancelled; new claims 15, 16 and 17 added; all other claims unchanged."
4. [Where various kinds of amendments are made]:
"Claims 1-10 unchanged; claims 11 to 13, 18 and 19 cancelled; claims 14, 15 and 16 replaced by amended claim 14; claim 17 subdivided into amended claims 15, 16 and 17; new claims 20 and 21 added."

"Statement under Article 19(1)" (Rule 46.4)

The amendments may be accompanied by a statement explaining the amendments and indicating any impact that such amendments might have on the description and the drawings (which cannot be amended under Article 19(1)).

The statement will be published with the international application and the amended claims.

It must be in the language in which the international application is to be published.

It must be brief, not exceeding 500 words if in English or if translated into English.

It should not be confused with and does not replace the letter indicating the differences between the claims as filed and as amended. It must be filed on a separate sheet and must be identified as such by a heading, preferably by using the words "Statement under Article 19(1)."

It may not contain any disparaging comments on the international search report or the relevance of citations contained in that report. Reference to citations, relevant to a given claim, contained in the international search report may be made only in connection with an amendment of that claim.

Consequence if a demand for international preliminary examination has already been filed

If, at the time of filing any amendments and any accompanying statement, under Article 19, a demand for international preliminary examination has already been submitted, the applicant must preferably, at the time of filing the amendments (and any statement) with the International Bureau, also file with the International Preliminary Examining Authority a copy of such amendments (and of any statement) and, where required, a translation of such amendments for the procedure before that Authority (see Rules 55.3(a) and 62.2, first sentence). For further information, see the Notes to the demand form (PCT/IPEA/401).

Consequence with regard to translation of the international application for entry into the national phase

The applicant's attention is drawn to the fact that, upon entry into the national phase, a translation of the claims as amended under Article 19 may have to be furnished to the designated/elected Offices, instead of, or in addition to, the translation of the claims as filed.

For further details on the requirements of each designated/elected Office, see the *PCT Applicant's Guide*, Volume II.

ITEM 6

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
3 June 2004 (03.06.2004)

PCT

(10) International Publication Number
WO 2004/045543 A2

(51) International Patent Classification⁷: **A61K**
(21) International Application Number:
PCT/US2003/036787
(22) International Filing Date:
14 November 2003 (14.11.2003)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/426,137 14 November 2002 (14.11.2002) US
60/502,050 10 September 2003 (10.09.2003) US

(71) Applicant: **DHARMACON, INC.** [US/US]; 2650 Crescent Drive, Suite 100, Lafayette, CO 80026 (US).

(72) Inventors: **ANASTASIA, Khvorova**; 635 Walden #204, Denver, CO (US). **ANGELA, Reynolds**; 11445 Conifer Ridge Drive, Conifer, CO 80433 (US). **DEVIN, Leake**; 3050 Krameria Street, Denver, CO 80303 (US). **WILLIAM, Marshall**; 495 Mohawk Drive, Denver, CO 80303 (US). **Stephen, Scaringe**; 2746 Pairie Ridge, Lafayette, CO 80026 (US).

(74) Agent: **KALOW, David**; Kalow & Springut LLP, 488 Madison Ave, 19th Floor, New York, NY 10022 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

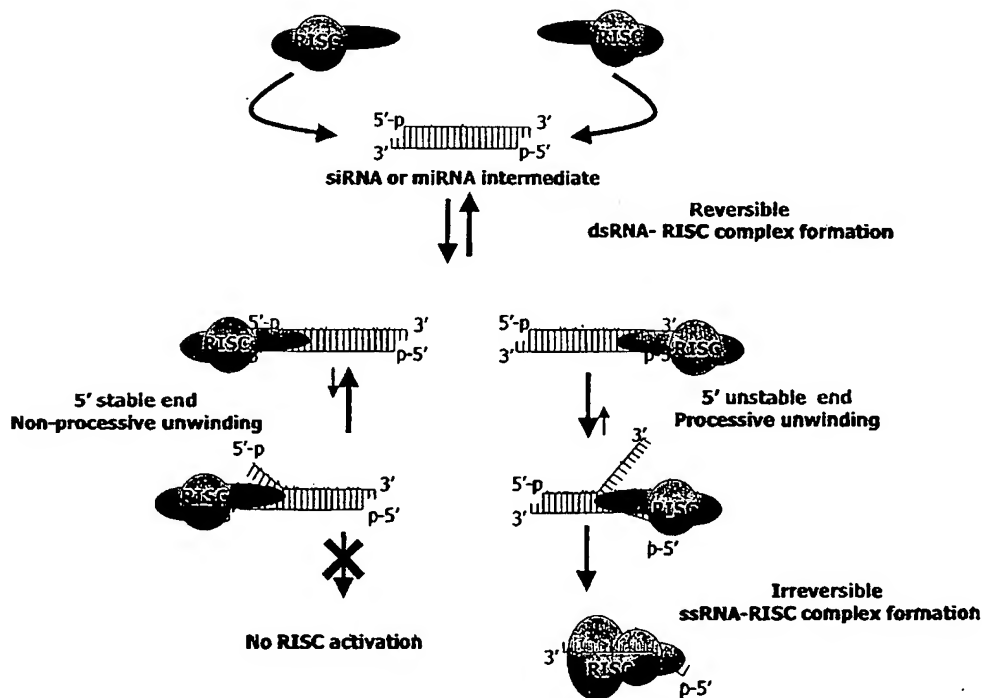
(84) Designated States (*regional*): ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

[Continued on next page]

(54) Title: FUNCTIONAL AND HYPERFUNCTIONAL siRNA



(57) Abstract: Efficient sequence specific gene silencing is possible through the use of siRNA technology. By selecting particular siRNAs by rationale design, one can maximize the generation of an effective gene silencing reagent, as well as methods for silencing genes.

WO 2004/045543 A2

Functional and Hyperfunctional siRNA

5 **Cross Reference to Related Applications**

 This application claims the benefit of the filing date of U.S. Provisional Application Serial No. 60/426,137, filed November 14, 2002, entitled "Combinatorial Pooling Approach for siRNA Induced Gene Silencing and Methods for Selecting siRNA," and U.S. Provisional Application Serial No. 60/502,050, filed September 10, 10 2003, entitled "Methods for Selecting siRNA," the entire disclosures of which are hereby incorporated by reference into the present disclosure.

Field of Invention

 The present invention relates to RNA interference ("RNAi").

15

Background of the Invention

 Relatively recently, researchers observed that double stranded RNA ("dsRNA") could be used to inhibit protein expression. This ability to silence a gene has broad potential for treating human diseases, and many researchers and 20 commercial entities are currently investing considerable resources in developing therapies based on this technology.

 Double stranded RNA induced gene silencing can occur on at least three different levels: (i) transcription inactivation, which refers to RNA guided DNA or 25 histone methylation; (ii) siRNA induced mRNA degradation; and (iii) mRNA induced transcriptional attenuation.

 It is generally considered that the major mechanism of RNA induced silencing (RNA interference, or RNAi) in mammalian cells is mRNA degradation. Initial 30 attempts to use RNAi in mammalian cells focused on the use of long strands of dsRNA. However, these attempts to induce RNAi met with limited success, due in part to the induction of the interferon response, which results in a general, as opposed to a target-specific, inhibition of protein synthesis. Thus, long dsRNA is not a viable option for RNAi in mammalian systems.

More recently it has been shown that when short (18-30 bp) RNA duplexes are introduced into mammalian cells in culture, sequence-specific inhibition of target mRNA can be realized without inducing an interferon response. Certain of these short dsRNAs, referred to as small inhibitory RNAs ("siRNAs"), can act catalytically at sub-molar concentrations to cleave greater than 95% of the target mRNA in the cell. A description of the mechanisms for siRNA activity, as well as some of its applications are described in Provost *et al.*, *Ribonuclease Activity and RNA Binding of Recombinant Human Dicer*, E.M.B.O. J., 2002 Nov. 1; 21(21): 5864–5874; Tabara *et al.*, *The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1 and a DexH-box Helicase to Direct RNAi in C. elegans*, Cell 2002, June 28;109(7):861-71; Ketting *et al.*, *Dicer Functions in RNA Interference and in Synthesis of Small RNA Involved in Developmental Timing in C. elegans*; Martinez *et al.*, *Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi*, Cell 2002, Sept. 6; 110(5):563; Hutvagner & Zamore, *A microRNA in a multiple-turnover RNAi enzyme complex*, Science 2002, 297:2056.

From a mechanistic perspective, introduction of long double stranded RNA into plants and invertebrate cells is broken down into siRNA by a Type III endonuclease known as Dicer. Sharp, *RNA interference—2001*, Genes Dev. 2001, 15:485. Dicer, a ribonuclease-III-like enzyme, processes the dsRNA into 19-23 base pair short interfering RNAs with characteristic two base 3' overhangs. Bernstein, Caudy, Hammond, & Hannon, *Role for a bidentate ribonuclease in the initiation step of RNA interference*, Nature 2001, 409:363. The siRNAs are then incorporated into an RNA-induced silencing complex (RISC) where one or more helicases unwind the siRNA duplex, enabling the complementary antisense strand to guide target recognition. Nykanen, Haley, & Zamore, *ATP requirements and small interfering RNA structure in the RNA interference pathway*, Cell 2001, 107:309. Upon binding to the appropriate target mRNA, one or more endonucleases within the RISC cleaves the target to induce silencing. Elbashir, Lendeckel, & Tuschl, *RNA interference is mediated by 21- and 22-nucleotide RNAs*, Genes Dev 2001, 15:188, **Figure 1.**

The interference effect can be long lasting and may be detectable after many cell divisions. Moreover, RNAi exhibits sequence specificity. Kisielow, M. *et al.*

(2002) *Isoform-specific knockdown and expression of adaptor protein ShcA using small interfering RNA*, J. of Biochemistry 363: 1-5. Thus, the RNAi machinery can specifically knock down one type of transcript, while not affecting closely related mRNA. These properties make siRNA a potentially valuable tool for inhibiting gene expression and studying gene function and drug target validation. Moreover, siRNAs are potentially useful as therapeutic agents against: (1) diseases that are caused by over-expression or misexpression of genes; and (2) diseases brought about by expression of genes that contain mutations.

10 Successful siRNA-dependent gene silencing depends on a number of factors. One of the most contentious issues in RNAi is the question of the necessity of siRNA design, *i.e.*, considering the sequence of the siRNA used. Early work in *C. elegans* and plants circumvented the issue of design by introducing long dsRNA (see, for instance, Fire, A. *et al.* (1998) *Nature* 391:806-811). In this primitive organism, long
15 dsRNA molecules are cleaved into siRNA by Dicer, thus generating a diverse population of duplexes that can potentially cover the entire transcript. While some fraction of these molecules are non-functional (*i.e.* induce little or no silencing) one or more have the potential to be highly functional, thereby silencing the gene of interest and alleviating the need for siRNA design. Unfortunately, due to the interferon
20 response, this same approach is unavailable for mammalian systems. While this effect can be circumvented by bypassing the Dicer cleavage step and directly introducing siRNA, this tactic carries with it the risk that the chosen siRNA sequence may be non-functional or semi-functional.

25 A number of researches have expressed the view that siRNA design is not a crucial element of RNAi. On the other hand, others in the field have begun to explore the possibility that RNAi can be made more efficient by paying attention to the design, of the siRNA. Unfortunately, none of the reported methods have provided a satisfactory scheme for reliably selecting siRNA with acceptable levels of
30 functionality. Accordingly, there is a need to develop rational criteria by which to select siRNA with an acceptable level of functionality, and to identify siRNA that have this improved level of functionality, as well as to identify siRNAs that are hyperfunctional.

+

5

10

Formula I

Formula II

Formula III

25 Formula IV

30 Formula V

Formula VI

Relative functionality of siRNA = $-(G_{13})^3 - (C_{19}) + (A_{19})^2 + (A_3)$

Formula VII

Relative functionality of siRNA = $-(GC/2) + (AU_{15-19})/2 - (T_{m20^\circ C})^*1 - (G_{13})^3 - (C_{19})$
 5 $+(A_{19})^3 + (A_3)^3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2$

wherein in Formulas I – VII:

$T_{m\ 20^\circ C} = 1$ if the T_m is greater than $20^\circ C$;

10 $A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at

positions 15 – 19;

15 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

20 $A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

25 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

or

30 Formula VIII Relative functionality of siRNA =

$$\begin{aligned} & (-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6- \\ & 9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19} \\ & +7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+ \end{aligned}$$

$$11 * C_9 + 15 * G_1 + 18 * A_3 + 19 * U_{10} - T_m - 3 * (GC_{total}) - 6 * (GC_{15-19}) - 30 * X; \text{ and}$$

Formula IX Relative functionality of siRNA =

$$\begin{aligned} & (14.1) * A_3 + (14.9) * A_6 + (17.6) * A_{13} + (24.7) * A_{19} + (14.2) * U_{10} + (10.5) * \\ & C_9 + (23.9) * G_1 + (16.3) * G_2 + (-12.3) * A_{11} + (-19.3) * U_1 + (-12.1) * U_2 + \\ & (-11) * U_3 + (-15.2) * U_{15} + (-11.3) * U_{16} + (-11.8) * C_3 + (-17.4) * C_6 + (- \\ & 10.5) * C_7 + (-13.7) * G_{13} + (-25.9) * G_{19} - T_m - 3 * (GC_{total}) - 6 * (GC_{15-19}) - \\ & 30 * X \end{aligned}$$

10 wherein

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;

$A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;

$A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;

$A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;

15 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;

$A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;

$A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;

$A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;

20 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;

25 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;

$C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;

$C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;

$C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;

$C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;

30 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;

$C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
5 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is
present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
10 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
15 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

20 GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand
or within positions 15 –18 if the sense strand is only 18 base pairs in length;
 GC_{total} = the number of G and C bases in the sense strand;
 $T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs,
25 otherwise its value is 0; and

X = the number of times that the same nucleotide repeats four or more times in a row.

According to a third embodiment, the present invention is directed to a kit
comprised of at least one siRNA that contains a sequence that is optimized according
30 to one of the formulas above. Preferably the kit contains at least two optimized
siRNA, each of which comprises a duplex, wherein one strand of each duplex
comprises at least eighteen contiguous bases that are complementary to a region of a
target messenger RNA. For mammalian systems, the siRNA preferably comprises
between 18 and 30 nucleotide base pairs.

The ability to use the above algorithms, which are not sequence or species specific, allows for the cost-effective selection of optimized siRNAs for specific target sequences. Accordingly, there will be both greater efficiency and reliability in the use of siRNA technologies.

According to a fourth embodiment, the present invention provides a method for developing an siRNA algorithm for selecting functional and hyperfunctional siRNAs for a given sequence. The method comprises:

- (a) selecting a set of siRNAs;
- (b) measuring the gene silencing ability of each siRNA from said set;
- (c) determining the relative functionality of each siRNA;
- (d) determining the amount of improved functionality by the presence or absence of at least one variable selected from the group consisting of the total GC content, melting temperature of the siRNA, GC content at positions 15 –19, the presence or absence of a particular nucleotide at a particular position and the number of times that the same nucleotide repeats within a given sequence; and
- (e) developing an algorithm using the information of step (d).

According to this embodiment, preferably the set of siRNAs comprises at least 90 siRNAs from at least one gene, more preferably at least 180 siRNAs from at least two different genes, and most preferably at least 270 and 360 siRNAs from at least three and four different genes, respectively. Additionally, in step (d) the determination is made with preferably at least two, more preferably at least three, even more preferably at least four, and most preferably all of the variables. The resulting algorithm is not target sequence specific.

In a fifth embodiment, the present invention provides rationally designed siRNAs identified using the formulas above.

In a sixth embodiment, the present invention is directed to hyperfunctional siRNA.

For a better understanding of the present invention together with other and further advantages and embodiments, reference is made to the following description taken in conjunction with the examples, the scope of which is set forth in the appended claims.

5

Brief Description of the Figures

Figure 1 shows a model for siRNA-RISC interactions. RISC has the ability to interact with either end of the siRNA or miRNA molecule. Following binding, the duplex is unwound, and the relevant target is identified, cleaved, and released.

10

Figure 2 is a representation of the functionality of two hundred and seventy siRNA duplexes that were generated to target human cyclophilin, human diazepam-binding inhibitor (DBI), and firefly luciferase.

15 **Figure 3a** is a representation of the silencing effect of 30 siRNAs in three different cells lines, HEK293, DU145, and Hela. **Figure 3b** shows the frequency of different functional groups (>95% silencing (black), >80% silencing (gray), >50% silencing (dark gray), and <50% silencing (white)) based on GC content. In cases where a given bar is absent from a particular GC percentage, no siRNA were identified for that particular group. **Figure 3c** shows the frequency of different functional groups based on melting temperature (T_m). Again, each group has four different divisions: >95% (black), >80% (gray), >50% (dark gray), and <50% (white) silencing.

25 **Figure 4** is a representation of a statistical analysis that revealed correlations between silencing and five sequence-related properties of siRNA: (A) an A at position 19 of the sense strand, (B) an A at position 3 of the sense strand, (C) a U at position 10 of the sense strand, (D) a base other than G at position 13 of the sense strand, and (E) a base other than C at position 19 of the sense strand. All variables were correlated with siRNA silencing of firefly luciferase and human cyclophilin. SiRNAs satisfying the criterion are grouped on the left (Selected) while those that do not, are grouped on the right (Eliminated). Y-axis is "% Silencing of Control." Each position on the X-axis represents a unique siRNA.

30

Figures 5 A and 5 B are representations of firefly luciferase and cyclophilin siRNA panels sorted according to functionality and predicted values using Formula VIII. The siRNA found within the circle represent those that have Formula VIII values (SMARTscores™) above zero. SiRNA outside the indicated area have calculated Formula VIII values that are below zero. Y-axis is “Expression (% Control).” Each position on the X-axis represents a unique siRNA.

Figure 6A is a representation of the average internal stability profile (AISP) derived from 270 siRNAs taken from three separate genes (cyclophilin B, DBI and firefly luciferase). Graphs represent AISP values of highly functional, functional, and non-functional siRNA. **Figure 6B** is a comparison between the AISP of naturally derived GFP siRNA (filled squares) and the AISP of siRNA from cyclophilin B, DBI, and luciferase having >90% silencing properties (no fill) for the antisense strand. “DG” is the symbol for ΔG , free energy.

15

Figure 7 is a histogram showing the differences in duplex functionality upon introduction of basepair mismatches. The X-axis shows the mismatch introduced in the siRNA and the position it is introduced (*e.g.*, 8C->A reveals that position 8 (which normally has a C) has been changed to an A). The Y-axis is “% Silencing (Normalized to Control).”

20

Figure 8a is histogram that shows the effects of 5' sense and antisense strand modification with 2'-O-methylation on functionality. **Figure 8b** is an expression profile showing a comparison of sense strand off-target effects for IGF1R-3 and 2'-O-methyl IGF1R-3. Sense strand off-targets (lower white box) are not induced when the 5' end of the sense strand is modified with 2'-O-methyl groups (top white box).

25

Figure 9 shows a graph of SMARTscores™ versus RNAi silencing values for more than 360 siRNA directed against 30 different genes. SiRNA to the right of the vertical bar represent those siRNA that have desirable SMARTscores™.

30

Figures 10A – E compare the RNAi of five different genes (SEAP, DBI, PLK, Firefly Luciferase, and Renilla Luciferase) by varying numbers of randomly selected

siRNA and four rationally designed (SMART-selected) siRNA chosen using the algorithm described in Formula VIII. In addition, RNAi induced by a pool of the four SMART-selected siRNA is reported at two different concentrations (100 and 400nM). **10F** is a comparison between a pool of randomly selected EGFR siRNA (Pool 1) and a pool of SMART selected EGFR siRNA (Pool 2). Pool 1, S1—S4 and Pool 2 S1—S4 represent the individual members that made up each respective pool. Note that numbers for random siRNAs represent the position of the 5' end of the sense strand of the duplex. The Y-axis represents the % expression of the control(s). The X-axis is the percent expression of the control.

10

Figure 11 shows the Western blot results from cells treated with siRNA directed against twelve different genes involved in the clathrin-dependent endocytosis pathway (CHC, DynII, CALM, CLCa, CLCb, Eps15, Eps15R, Rab5a, Rab5b, Rab5c, β 2 subunit of AP-2 and EEA.1). SiRNA were selected using Formula VIII. “Pool” represents a mixture of duplexes 1-4. Total concentration of each siRNA in the pool is 25 nM. Total concentration = $4 \times 25 = 100$ nM.

Figure 12 is a representation of the gene silencing capabilities of rationally-selected siRNA directed against ten different genes (human and mouse cyclophilin, C-myc, human lamin A/C, QB (ubiquinol-cytochrome c reductase core protein I), MEK1 and MEK2, ATE1 (arginyl-tRNA protein transferase), GAPDH, and Eg5). The Y-axis is the percent expression of the control. Numbers 1, 2, 3 and 4 represent individual rationally selected siRNA. “Pool” represents a mixture of the four individual siRNA.

Figure 13 is the sequence of the top ten Bcl2 siRNAs as determined by Formula VIII. Sequences are listed 5' to 3'.

Figure 14 is the knockdown by the top ten Bcl2 siRNAs at 100nM concentrations. The Y-axis represents the amount of expression relative to the non-specific (ns) and transfection mixture control.

Figure 15 represents a functional walk where siRNA beginning on every other base pair of a region of the luciferase gene are tested for the ability to silence the luciferase

gene. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of each individual siRNA.

Figure 16 is a histogram demonstrating the inhibition of target gene expression by pools of 2 and 3 siRNAs duplexes taken from the walk described in Figure 15. The Y-axis represents the percent expression relative to control. The X-axis represents the position of the first siRNA in paired pools, or trios of siRNA. For instance, the first paired pool contains siRNA 1 and 3. The second paired pool contains siRNA 3 and 5. Pool 3 (of paired pools) contains siRNA 5 and 7, and so on.

10

Figure 17 is a histogram demonstrating the inhibition of target gene expression by pools of 4 and 5 siRNA duplexes. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of the first siRNA in each pool.

15

Figure 18 is a histogram demonstrating the inhibition of target gene expression by siRNAs that are ten and twenty basepairs apart. The Y-axis represents the percent expression relative to a control. The X-axis represents the position of the first siRNA in each pool.

20

Figure 19 shows that pools of siRNAs (dark gray bar) work as well (or better) than the best siRNA in the pool (light gray bar). The Y-axis represents the percent expression relative to a control. The X-axis represents the position of the first siRNA in each pool.

25

Figure 20 shows that the combination of several semifunctional siRNAs (dark gray) result in a significant improvement of gene expression inhibition over individual (semi-functional; light gray) siRNA. The Y-axis represents the percent expression relative to a control.

30

Figure 21 shows both pools (Library, Lib) and individual siRNAs in inhibition of gene expression of Beta-Galactosidase, Renilla Luciferase and SEAP (alkaline phosphatase). Numbers on the X-axis indicate the position of the 5'-most nucleotide of the sense strand of the duplex. The Y-axis represents the percent expression of

each gene relative to a control. Libraries contain siRNAs that begin at the following nucleotides: Seap: Lib 1: 206, 766, 812,923, Lib 2: 1117, 1280, 1300, 1487, Lib 3: 206, 766, 812, 923, 1117, 1280, 1300,1487, Lib 4: 206, 812, 1117, 1300, Lib 5: 766, 923, 1280, 1487, Lib 6: 206, 1487; Bgal: Lib 1: 979, 1339, 2029, 2590, Lib 2: 1087,1783,2399,3257, Lib 3: 979, 1783, 2590, 3257, Lib 4: 979, 1087, 1339, 1783, 2029, 2399,2590,3257, Lib 5: 979, 1087, 1339, 1783, Lib 6: 2029,2399,2590,3257; Renilla: Lib 1: 174,300,432,568, Lib 2: 592, 633, 729,867, Lib 3: 174, 300, 432, 568, 592, 633,729,867, Lib 4: 174, 432, 592, 729, Lib 5: 300,568,633,867, Lib 6: 592,568.

10

Figure 22 showS the results of an EGFR and TfnR internalization assay when single gene knockdowns are performed. The Y-axis represents percent internalization relative to control.

15

Figure 23 shows the results of an EGFR and TfnR internalization assay when multiple genes are knocked down (*e.g.* Rab5a, b, c). The Y-axis represents the percent internalization relative to control.

Figure 24 shows the simultaneous knockdown of four different genes. SiRNAs directed against G6PD, GAPDH, PLK, and UBQ were simultaneously introduced into cells. Twenty-four hours later, cultures were harvested and assayed for mRNA target levels for each of the four genes. A comparison is made between cells transfected with individual siRNAs vs. a pool of siRNAs directed against all four genes.

25

Figure 25 shows the functionality of ten siRNAs at 0.3nM concentrations.

Detailed Description

Definitions

Unless stated otherwise, the following terms and phrases have the meanings provided below:

30

siRNA

The term “siRNA” refers to small inhibitory RNA duplexes that induce the RNA interference (RNAi) pathway. These molecules can vary in length (generally between 18-30 basepairs) and contain varying degrees of complementarity to their target mRNA in the antisense strand. Some, but not all, siRNA have unpaired overhanging bases on the 5’ or 3’ end of the sense strand and/or the antisense strand. The term “siRNA” includes duplexes of two separate strands, as well as single strands that can form hairpin structures comprising a duplex region.

SiRNA may be divided into five (5) groups (non-functional, semi-functional, functional, highly functional, and hyper-functional) based on the level or degree of silencing that they induce in cultured cell lines. As used herein, these definitions are based on a set of conditions where the siRNA is transfected into said cell line at a concentration of 100nM and the level of silencing is tested at a time of roughly 24 hours after transfection, and not exceeding 72 hours after transfection. In this context, “non-functional siRNA” are defined as those siRNA that induce less than 50% (<50%) target silencing. “Semi-functional siRNA” induce 50-79% target silencing. “Functional siRNA” are molecules that induce 80-95% gene silencing. “Highly-functional siRNA” are molecules that induce greater than 95% gene silencing. “Hyperfunctional siRNA” are a special class of molecules: For purposes of this document, hyperfunctional siRNA are defined as those molecules that: (1) induce greater than 95% silencing of a specific target when they are transfected at subnanomolar concentrations (*i.e.*, less than one nanomolar); and/or (2) induce functional (or better) levels of silencing for greater than 96 hours. These relative functionalities (though not intended to be absolutes) may be used to compare siRNAs to a particular target for applications such as functional genomics, target identification and therapeutics.

miRNA

The term “miRNA” refers to microRNA.

Gene silencing

The phrase “gene silencing” refers to a process by which the expression of a specific gene product is lessened or attenuated. Gene silencing can take place by a

variety of pathways. Unless specified otherwise, as used herein, gene silencing refers to decreases in gene product expression that results from RNA interference (RNAi), a defined, though partially characterized pathway whereby small inhibitory RNA (siRNA) act in concert with host proteins (e.g. the RNA induced silencing complex, RISC) to degrade messenger RNA (mRNA) in a sequence-dependent fashion. The level of gene silencing can be measured by a variety of means, including, but not limited to, measurement of transcript levels by Northern Blot Analysis, B-DNA techniques, transcription-sensitive reporter constructs, expression profiling (e.g. DNA chips), and related technologies. Alternatively, the level of silencing can be measured by assessing the level of the protein encoded by a specific gene. This can be accomplished by performing a number of studies including Western Analysis, measuring the levels of expression of a reporter protein that has e.g. fluorescent properties (e.g. GFP) or enzymatic activity (e.g. alkaline phosphatases), or several other procedures.

Transfection

The term “transfection” refers to a process by which agents are introduced into a cell. The list of agents that can be transfected is large and includes, but is not limited to, siRNA, sense and/or anti-sense sequences, DNA encoding one or more genes and organized into an expression plasmid, proteins, protein fragments, and more. There are multiple methods for transfecting agents into a cell including, but not limited to, electroporation, calcium phosphate-based transfections, DEAE-dextran-based transfections, lipid-based transfections, molecular conjugate-based transfections (e.g. polylysine-DNA conjugates), microinjection and others.

Target

The term “target” is used in a variety of different forms throughout this document and is defined by the context in which it is used. “Target mRNA” refers to a messenger RNA to which a given siRNA can be directed against. “Target sequence” and “target site” refer to a sequence within the mRNA to which the sense strand of an siRNA shows varying degrees of homology and the antisense strand exhibits varying degrees of complementarity. The term “siRNA target” can refer to the gene, mRNA, or protein against which an siRNA is directed. Similarly “target silencing” can refer to the state of a gene, or the corresponding mRNA or protein.

Off-target silencing and Off-target interference

The phrases “off-target silencing” and “off-target interference” are defined as degradation of mRNA other than the intended target mRNA due to overlapping and/or partial homology with secondary mRNA messages.

SMARTscore™

The term “SMARTscore™” refers to a number determined by applying any of the Formulas I - Formula IX to a given siRNA sequence. The term “SMART-selected” or “rationally selected” or “rational selection” refers to siRNA that have been selected on the basis of their SMARTscores™.

Complementary

The term “complementary” refers to the ability of polynucleotides to form base pairs with one another. Base pairs are typically formed by hydrogen bonds between nucleotide units in antiparallel polynucleotide strands. Complementary polynucleotide strands can base pair in the Watson-Crick manner (*e.g.*, A to T, A to U, C to G), or in any other manner that allows for the formation of duplexes. As persons skilled in the art are aware, when using RNA as opposed to DNA, uracil rather than thymine is the base that is considered to be complementary to adenosine. However, when a U is denoted in the context of the present invention, the ability to substitute a T is implied, unless otherwise stated.

Perfect complementarity or 100% complementarity refers to the situation in which each nucleotide unit of one polynucleotide strand can hydrogen bond with a nucleotide unit of a second polynucleotide strand. Less than perfect complementarity refers to the situation in which some, but not all, nucleotide units of two strands can hydrogen bond with each other. For example, for two 20-mers, if only two base pairs on each strand can hydrogen bond with each other, the polynucleotide strands exhibit 10% complementarity. In the same example, if 18 base pairs on each strand can hydrogen bond with each other, the polynucleotide strands exhibit 90% complementarity. “Substantial complementarity” refers to polynucleotide strands exhibiting 79% or greater complementarity, excluding regions of the polynucleotide strands, such as overhangs, that are selected so as to be noncomplementary.

(“Substantial similarity” refers to polynucleotide strands exhibiting 79% or greater similarity, excluding regions of the polynucleotide strands, such as overhangs, that are selected so as not to be similar.) Thus, for example, two polynucleotides of 29 nucleotide units each, wherein each comprises a di-dT at the 3’ terminus such that the duplex region spans 27 bases, and wherein 26 of the 27 bases of the duplex region on each strand are complementary, are substantially complementary since they are 96.3% complementary when excluding the di-dT overhangs.

Deoxynucleotide

The term “deoxynucleotide” refers to a nucleotide or polynucleotide lacking a hydroxyl group (OH group) at the 2’ and/or 3’ position of a sugar moiety. Instead, it has a hydrogen bonded to the 2’ and/or 3’ carbon. Within an RNA molecule that comprises one or more deoxynucleotides, “deoxynucleotide” refers to the lack of an OH group at the 2’ position of the sugar moiety, having instead a hydrogen bonded directly to the 2’ carbon.

Deoxyribonucleotide

The terms “deoxyribonucleotide” and “DNA” refer to a nucleotide or polynucleotide comprising at least one sugar moiety that has an H, rather than an OH, at its 2’ and/or 3’ position.

Substantially Similar

The phrase “substantially similar” refers to a similarity of at least 90% with respect to the identity of the bases of the sequence.

Duplex Region

The phrase “duplex region” refers to the region in two complementary or substantially complementary polynucleotides that form base pairs with one another, either by Watson-Crick base pairing or any other manner that allows for a stabilized duplex between polynucleotide strands that are complementary or substantially complementary. For example, a polynucleotide strand having 21 nucleotide units can base pair with another polynucleotide of 21 nucleotide units, yet only 19 bases on each strand are complementary or substantially complementary, such that the “duplex region” has 19 base pairs. The remaining bases may, for example, exist as 5’ and 3’

overhangs. Further, within the duplex region, 100% complementarity is not required; substantial complementarity is allowable within a duplex region. Substantial complementarity refers to 79% or greater complementarity. For example, a mismatch in a duplex region consisting of 19 base pairs results in 94.7% complementarity, rendering the duplex region substantially complementary.

Nucleotide

The term "nucleotide" refers to a ribonucleotide or a deoxyribonucleotide or modified form thereof, as well as an analog thereof. Nucleotides include species that comprise purines, *e.g.*, adenine, hypoxanthine, guanine, and their derivatives and analogs, as well as pyrimidines, *e.g.*, cytosine, uracil, thymine, and their derivatives and analogs.

Nucleotide analogs include nucleotides having modifications in the chemical structure of the base, sugar and/or phosphate, including, but not limited to, 5-position pyrimidine modifications, 8-position purine modifications, modifications at cytosine exocyclic amines, and substitution of 5-bromo-uracil; and 2'-position sugar modifications, including but not limited to, sugar-modified ribonucleotides in which the 2'-OH is replaced by a group such as an H, OR, R, halo, SH, SR, NH₂, NHR, NR₂, or CN, wherein R is an alkyl moiety. Nucleotide analogs are also meant to include nucleotides with bases such as inosine, queuosine, xanthine, sugars such as 2'-methyl ribose, non-natural phosphodiester linkages such as methylphosphonates, phosphorothioates and-peptides.

Modified bases refer to nucleotide bases such as, for example, adenine, guanine, cytosine, thymine, uracil, xanthine, inosine, and queuosine that have been modified by the replacement or addition of one or more atoms or groups. Some examples of types of modifications that can comprise nucleotides that are modified with respect to the base moieties include but are not limited to, alkylated, halogenated, thiolated, aminated, amidated, or acetylated bases, individually or in combination. More specific examples include, for example, 5-propynyluridine, 5-propynylcytidine, 6-methyladenine, 6-methylguanine, N,N,-dimethyladenine, 2-propyladenine, 2-propylguanine, 2-aminoadenine, 1-methylinosine, 3-methyluridine, 5-methylcytidine, 5-methyluridine and other nucleotides having a modification at the 5 position, 5-(2-

amino)propyl uridine, 5-halocytidine, 5-halouridine, 4-acetylcytidine, 1-methyladenosine, 2-methyladenosine, 3-methylcytidine, 6-methyluridine, 2-methylguanosine, 7-methylguanosine, 2,2-dimethylguanosine, 5-methylaminoethyluridine, 5-methyloxyuridine, deazanucleotides such as 7-deaza-adenosine, 6-azouridine, 6-azocytidine, 6-azothymidine, 5-methyl-2-thiouridine, other thio bases such as 2-thiouridine and 4-thiouridine and 2-thiocytidine, dihydrouridine, pseudouridine, queuosine, archaeosine, naphthyl and substituted naphthyl groups, any O- and N-alkylated purines and pyrimidines such as N6-methyladenosine, 5-methylcarbonylmethyluridine, uridine 5-oxyacetic acid, pyridine-4-one, pyridine-2-one, phenyl and modified phenyl groups such as aminophenol or 2,4,6-trimethoxy benzene, modified cytosines that act as G-clamp nucleotides, 8-substituted adenines and guanines, 5-substituted uracils and thymines, azapyrimidines, carboxyhydroxyalkyl nucleotides, carboxyalkylaminoalkyl nucleotides, and alkylcarbonylalkylated nucleotides. Modified nucleotides also include those nucleotides that are modified with respect to the sugar moiety, as well as nucleotides having sugars or analogs thereof that are not ribosyl. For example, the sugar moieties may be, or be based on, mannoses, arabinoses, glucopyranoses, galactopyranoses, 4'-thioribose, and other sugars, heterocycles, or carbocycles.

The term nucleotide is also meant to include what are known in the art as universal bases. By way of example, universal bases include but are not limited to 3-nitropyrrole, 5-nitroindole, or nebularine. The term "nucleotide" is also meant to include the N3' to P5' phosphoramidate, resulting from the substitution of a ribosyl 3' oxygen with an amine group.

25

Further, the term nucleotide also includes those species that have a detectable label, such as for example a radioactive or fluorescent moiety, or mass label attached to the nucleotide.

30 Polynucleotide

The term "polynucleotide" refers to polymers of nucleotides, and includes but is not limited to DNA, RNA, DNA/RNA hybrids including polynucleotide chains of regularly and/or irregularly alternating deoxyribosyl moieties and ribosyl moieties (*i.e.*, wherein alternate nucleotide units have an -OH, then and -H, then an -OH, then

an -H, and so on at the 2' position of a sugar moiety), and modifications of these kinds of polynucleotides, wherein the attachment of various entities or moieties to the nucleotide units at any position are included.

5 Polyribonucleotide

The term "polyribonucleotide" refers to a polynucleotide comprising two or more modified or unmodified ribonucleotides and/or their analogs. The term "polyribonucleotide" is used interchangeably with the term "oligoribonucleotide."

10 Ribonucleotide and ribonucleic acid

The term "ribonucleotide" and the phrase "ribonucleic acid" (RNA), refer to a modified or unmodified nucleotide or polynucleotide comprising at least one ribonucleotide unit. A ribonucleotide unit comprises an hydroxyl group attached to the 2' position of a ribosyl moiety that has a nitrogenous base attached in N-
15 glycosidic linkage at the 1' position of a ribosyl moiety, and a moiety that either allows for linkage to another nucleotide or precludes linkage.

Detailed Description of the Invention

20 The present invention is directed to improving the efficiency of gene silencing by siRNA. Through the inclusion of multiple siRNA sequences that are targeted to a particular gene and/or selecting an siRNA sequence based on certain defined criteria, improved efficiency may be achieved.

25 The present invention will now be described in connection with preferred embodiments. These embodiments are presented in order to aid in an understanding of the present invention and are not intended, and should not be construed, to limit the invention in any way. All alternatives, modifications and equivalents that may become apparent to those of ordinary skill upon reading this disclosure are included
30 within the spirit and scope of the present invention.

Furthermore, this disclosure is not a primer on RNA interference. Basic concepts known to persons skilled in the art have not been set forth in detail.

Optimizing siRNA

According to one embodiment, the present invention provides a method for improving the effectiveness of gene silencing for use to silence a particular gene through the selection of an optimal siRNA. An siRNA selected according to this method may be used individually, or in conjunction with the first embodiment, *i.e.*, with one or more other siRNAs, each of which may or may not be selected by this criteria in order to maximize their efficiency.

The degree to which it is possible to select an siRNA for a given mRNA that maximizes these criteria will depend on the sequence of the mRNA itself. However, the selection criteria will be independent of the target sequence. According to this method, an siRNA is selected for a given gene by using a rational design. That said, rational design can be described in a variety of ways. Rational design is, in simplest terms, the application of a proven set of criteria that enhance the probability of identifying a functional or hyperfunctional siRNA. In one method, rationally designed siRNA can be identified by maximizing one or more of the following criteria:

1. A low GC content, preferably between about 30–52%.
2. At least 2, preferably at least 3 A or U bases at positions 15–19 of the siRNA on the sense strand.
3. An A base at position 19 of the sense strand.
4. An A base at position 3 of the sense strand.
5. A U base at position 10 of the sense strand.
6. An A base at position 14 of the sense strand.
7. A base other than C at position 19 of the sense strand.
8. A base other than G at position 13 of the sense strand.
9. A T_m , which refers to the character of the internal repeat that results in inter- or intramolecular structures for one strand of the duplex, that is preferably not stable at greater than 50°C, more preferably not stable at greater than 37°C, even more preferably not stable at greater than 30°C and most preferably not stable at greater than 20°C.
10. A base other than U at position 5 of the sense strand.
11. A base other than A at position 11 of the sense strand.

Criteria 5, 6, 10 and 11 are minor criteria, but are nonetheless desirable. Accordingly, preferably an siRNA will satisfy as many of the aforementioned criteria as possible, more preferably at least 1 – 4 and 7-9, and most preferably all of the
 5 criteria

With respect to the criteria, GC content, as well as a high number of AU in positions 15-19, may be important for easement of the unwinding of double stranded siRNA duplex. Duplex unwinding has been shown to be crucial for siRNA
 10 functionality *in vivo*.

With respect to criterion 9, the internal structure is measured in terms of the melting temperature of the single strand of siRNA, which is the temperature at which 50% of the molecules will become denatured. With respect to criteria 2 – 8 and 10 –
 15 11, the positions refer to sequence positions on the sense strand, which is the strand that is identical to the mRNA.

In one preferred embodiment, at least criteria 1 and 8 are satisfied. In another preferred embodiment, at least criteria 7 and 8 are satisfied. In still another preferred
 20 embodiment, at least criteria 1, 8 and 9 are satisfied.

It should be noted that all of the aforementioned criteria regarding sequence position specifics are with respect to the ~~the 5'~~ end of the sense strand. Reference is made to the sense strand, because most databases contain information that describes
 25 the information of the mRNA. Because according to the present invention a chain can be from 18 to 30 bases in length, and the aforementioned criteria assumes a chain 19 base pairs in length, it is important to keep the aforementioned criteria applicable to the correct bases.

30 When there are only 18 bases, the base pair that is not present is the base pair that is located at the 3' of the sense strand. When there are twenty to thirty bases present, then additional bases are added at the 5' end of the sense chain and occupy positions 1 to 11. Accordingly, with respect to SEQ. ID NO. 0001. NNANANNNNUCNAANNNA and SEQ. ID NO. 0028.

GUCNNANANNNNUCNAANNNA, both would have A at position 3, A at position 5, U at position 10, C at position 11, A and position 13, A and position 14 and A at position 19. However, SEQ. ID NO. 0028 would also have C at position -1, U at position -2 and G at position -3.

5

For a 19 base pair siRNA, an optimal sequence of one of the strands may be represented below, where N is any base, A, C, G, or U:

- SEQ. ID NO. 0001. NNANANNNNUCNAANNNA
 10 SEQ. ID NO. 0002. NNANANNNNUGNAANNNA
 SEQ. ID NO. 0003. NNANANNNNUUNAANNNA
 SEQ. ID NO. 0004. NNANANNNNUCNCANNNA
 SEQ. ID NO. 0005. NNANANNNNUGNCANNNA
 SEQ. ID NO. 0006. NNANANNNNUUNCANNNA
 15 SEQ. ID NO. 0007. NNANANNNNUCNUANNNA
 SEQ. ID NO. 0008. NNANANNNNUG NUANNNA
 SEQ. ID NO. 0009. NNANANNNNUUNUANNNA
 SEQ. ID NO. 0010. NNANCNNNNUCNAANNNA
 SEQ. ID NO. 0011. NNANCNNNNUGNAANNNA
 20 SEQ. ID NO. 0012. NNANCNNNNUUNAANNNA
 SEQ. ID NO. 0013. NNANCNNNNUCNCANNNA
 SEQ. ID NO. 0014. NNANCNNNNUGNCANNNA
 SEQ. ID NO. 0015. NNANCNNNNUUNCANNNA
 SEQ. ID NO. 0016. NANCNNNNUCNUANNNA
 25 SEQ. ID NO. 0017. NNANCNNNNUGNUANNNA
 SEQ. ID NO. 0018. NNANCNNNNUUNUANNNA
 SEQ. ID NO. 0019. NNANGNNNNUCNAANNNA
 SEQ. ID NO. 0020. NNANGNNNNUGNAANNNA
 SEQ. ID NO. 0021. NNANGNNNNUUNAANNNA
 30 SEQ. ID NO. 0022. NNANGNNNNUCNCANNNA
 SEQ. ID NO. 0023. NNANGNNNNUGNCANNNA
 SEQ. ID NO. 0024. NNANGNNNNUUNCANNNA
 SEQ. ID NO. 0025. NNANGNNNNUCNUANNNA
 SEQ. ID NO. 0026. NNANGNNNNUGNUANNNA

In one embodiment, the sequence used as an siRNA is selected by choosing the siRNA that score highest according to one of the following seven algorithms that are represented by Formulas I - VII:

Formula I

$$\text{Relative functionality of siRNA} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

Formula II

$$\text{Relative functionality of siRNA} = -(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$$

Formula III

$$\text{Relative functionality of siRNA} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3$$

Formula IV

$$\text{Relative functionality of siRNA} =$$

$$-GC/2 + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

Formula V

$$\text{Relative functionality of siRNA} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11})$$

Formula VI

$$\text{Relative functionality of siRNA} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$$

Formula VII

$$\text{Relative functionality of siRNA} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2$$

In Formulas I - VII:

wherein $A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0,

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at

5 positions 15 –19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

10 $GC =$ the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

15 $A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0; and

20 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0.

Formulas I –VII provide relative information regarding functionality. When the values for two sequences are compared for a given formula, the relative functionality is ascertained; a higher positive number indicates a greater functionality.

25 For example, in many applications a value of 5 or greater is beneficial.

Additionally, in many applications, more than one of these formulas would provide useful information as to the relative functionality of potential siRNA sequences. However, it is beneficial to have more than one type of formula, because

30 not every formula will be able to help to differentiate among potential siRNA sequences. For example, in particularly high GC mRNAs, formulas that take that parameter into account would not be useful and application of formulas that lack GC elements (e.g., formulas V and VI) might provide greater insights into duplex functionality. Similarly, formula II might be used in situations where hairpin

structures are not observed in duplexes, and formula IV might be applicable for sequences that have higher AU content. Thus, one may consider a particular sequence in light of more than one or even all of these algorithms to obtain the best differentiation among sequences. In some instances, application of a given algorithm may identify an unusually large number of potential siRNA sequences, and in those cases, it may be appropriate to re-analyze that sequence with a second algorithm that is, for instance, more stringent. Alternatively, it is conceivable that analysis of a sequence with a given formula yields no acceptable siRNA sequences (*i.e.* low SMARTscores™). In this instance, it may be appropriate to re-analyze that sequences with a second algorithm that is, for instance, less stringent. In still other instances, analysis of a single sequence with two separate formulas may give rise to conflicting results (*i.e.* one formula generates a set of siRNA with high SMARTscores™ while the other formula identifies a set of siRNA with low SMARTscores™). In these instances, it may be necessary to determine which weighted factor(s) (*e.g.* GC content) are contributing to the discrepancy and assessing the sequence to decide whether these factors should or should not be included. Alternatively, the sequence could be analyzed by a third, fourth, or fifth algorithm to identify a set of rationally designed siRNA.

The above-referenced criteria are particularly advantageous when used in combination with pooling techniques as depicted in Table I:

Table I

Criteria	Functional Probability					
	Oligos			Pools		
	>95%	>80%	<70%	>95%	>80%	<70%
Current	33.0	50.0	23.0	79.5	97.3	0.3
New	50.0	88.5	8.0	93.8	99.98	0.005
(GC)	28.0	58.9	36.0	72.8	97.1	1.6

The term “current” refers to Tuschl’s conventional siRNA parameters (Elbashir, S.M. et al. (2002) “Analysis of gene function in somatic mammalian cells using small interfering RNAs” *Methods* 26: 199-213). “New” refers to the design parameters

described in Formulas I-VII. "GC" refers to criteria that select siRNA solely on the basis of GC content.

As Table I indicates, when more functional siRNA duplexes are chosen,
 5 siRNAs that produce <70% silencing drops from 23% to 8% and the number of
 siRNA duplexes that produce >80% silencing rises from 50% to 88.5%. Further, of
 the siRNA duplexes with >80% silencing, a larger portion of these siRNAs actually
 silence >95% of the target expression (the new criteria increases the portion from
 33% to 50%). Using this new criteria in pooled siRNAs, shows that, with pooling, the
 10 amount of silencing >95% increases from 79.5% to 93.8% and essentially eliminates
 any siRNA pool from silencing less than 70%.

Table II similarly shows the particularly beneficial results of pooling in
 combination with the aforementioned criteria. However, Table II, which takes into
 15 account each of the aforementioned variables, demonstrates even a greater degree of
 improvement in functionality.

Table II

	Functional Probability					
	Oligos			Pools		
	Functional	Average	Non-functional	Functional	Average	Non-functional
Random	20	40	50	67	97	3
Criteria 1	52	99	0.1	97	93	0.0040
Criteria 4	89	99	0.1	99	99	0.0000

The terms "functional," "Average," and "Non-functional" refer to siRNA that exhibit
 20 >80%, >50%, and <50% functionality, respectively. Criteria 1 and 4 refer to specific
 criteria described above.

The above-described algorithms may be used with or without a computer
 program that allows for the inputting of the sequence of the mRNA and automatically
 25 outputs the optimal siRNA. The computer program may, for example, be accessible

from a local terminal or personal computer, over an internal network or over the Internet.

In addition to the formulas above, more detailed algorithms may be used for selecting siRNA. Preferably, at least one RNA duplex of between 18 and 30 base pairs is selected such that it is optimized according a formula selected from:

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-T_m-3*(GC_{total})-6*(GC_{15-19})-30*X$

wherein

$A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;

$A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;

$A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;

$A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;

$A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;

$A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;

$A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;

$A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;

$A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
5 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is
10 present or the sense strand is only 18 base pairs in length, its value is 0;

$G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
15 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is
present or the sense strand is only 18 base pairs in length, its value is 0;

20 $U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;
 $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
25 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
30 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

$GC_{15-19} =$ the number of G and C bases within positions 15 – 19 of the sense
strand, or within positions 15 –18 if the sense strand is only 18 base pairs in
length;

30

GC_{total} = the number of G and C bases in the sense strand;
T_m = 100 if the siRNA oligo has the internal repeat longer than 4 base pairs,
otherwise its value is 0; and
X = the number of times that the same nucleotide repeats four or more times in a
5 row.

The above formulas VIII and IX, as well as formulas I – VII, provide methods for selecting siRNA in order to increase the efficiency of gene silencing. A subset of variables of any of the formulas may be used, though when fewer variables are used,
10 the optimization hierarchy becomes less reliable.

With respect to the variables of the above-referenced formulas, a single letter of A or C or G or U followed by a subscript refers to a binary condition. The binary condition is that either the particular base is present at that particular position
15 (wherein the value is “1”) or the base is not present (wherein the value is “0”). Because position 19 is optional, *i.e.* there might be only 18 base pairs, when there are only 18 base pairs, any base with a subscript of 19 in the formulas above would have a zero value for that parameter. Before or after each variable is a number followed by
20 *, which indicates that the value of the variable is to be multiplied or weighed by that number.

The numbers preceding the variables A, or G, or C, or U in Formulas VIII and IX (or after the variables in Formula I - VII) were determined by comparing the difference in the frequency of individual bases at different positions in functional
25 siRNA and total siRNA. Specifically, the frequency in which a given base was observed at a particular position in functional groups was compared with the frequency that that same base was observed in the total, randomly selected siRNA set. If the absolute value of the difference between the functional and total values was found to be greater than 6%, that parameter was included in the equation. Thus for
30 instance, if the frequency of finding a “G” at position 13 (G₁₃) is found to be 6% in a given functional group, and the frequency of G₁₃ in the total population of siRNAs is 20%, the difference between the two values is 6%-20% = -14%. As the absolute value is greater than six (6), this factor (-14) is included in the equation. Thus in Formula VIII, in cases where the siRNA under study has a G in position 13, the accrued value

is $(-14) * (1) = -14$. In contrast, when a base other than G is found at position 13, the accrued value is $(-14) * (0) = 0$.

When developing a means to optimize siRNAs, the inventors observed that a
 5 bias toward low internal thermodynamic stability of the duplex at the 5'-antisense (AS) end is characteristic of naturally occurring miRNA precursors. The inventors extended this observation to siRNAs for which functionality had been assessed in tissue culture.

10 With respect to the parameter GC_{15-19} , a value of 0 – 5 will be ascribed depending on the number of G or C bases at positions 15 to 19. If there are only 18 base pairs, the value is between 0 and 4.

With respect to the criterion GC_{total} content, a number from 0 – 30 will be
 15 ascribed, which correlates to the total number of G and C nucleotides on the sense strand, excluding overhangs. Without wishing to be bound by any one theory, it is postulated that the significance of the GC content (as well as AU content at positions 15-19, which is a parameter for formulas III – VII) relates to the easement of the unwinding of a double-stranded siRNA duplex. Duplex unwinding is believed to be
 20 crucial for siRNA functionality *in vivo* and overall low internal stability, especially low internal stability of the first unwound base pair is believed to be important to maintain sufficient processivity of RISC complex-induced duplex unwinding. If the duplex has 19 base pairs, those at positions 15-19 on the sense strand will unwind first if the molecule exhibits a sufficiently low internal stability at that position. As
 25 persons skilled in the art are aware, RISC is a complex of approximately twelve proteins; Dicer is one, but not the only, helicase within this complex. Accordingly, although the GC parameters are believed to relate to activity with Dicer, they are also important for activity with other RISC proteins.

30 The value of the parameter T_m is 0 when there are no internal repeats longer than (or equal to) four base pairs present in the siRNA duplex; otherwise the value is 1. Thus for example, if the sequence ACGUACGU, or any other four nucleotide (or more) palindrome exists within the structure, the value will be one (1). Alternatively

if the structure ACGGACG, or any other 3 nucleotide (or less) palindrome exists, the value will be zero (0).

The variable "X" refers to the number of times that the same nucleotide occurs
 5 contiguously in a stretch of four or more units. If there are, for example, four
 contiguous As in one part of the sequence and elsewhere in the sequence four
 contiguous Cs, $X=2$. Further, if there are two separate contiguous stretches of four of
 the same nucleotides or eight or more of the same nucleotides in a row, then $X=2$.
 However, X does not increase for five, six or seven contiguous nucleotides.

10

Again, when applying Formula VIII or Formula IX to a given mRNA, (the
 "target RNA" or "target molecule"), one may use a computer program to evaluate the
 criteria for every sequence of 18 – 30 base pairs or only sequences of a fixed length,
 e.g., 19 base pairs. Preferably the computer program is designed such that it provides
 15 a report ranking of all of the potential siRNAs between 18 and 30 base pairs, ranked
 according to which sequences generate the highest value. A higher value refers to a
 more efficient siRNA for a particular target gene. The computer program that may be
 used, may be developed in any computer language that is known to be useful for
 scoring nucleotide sequences, or it may be developed with the assistance of
 20 commercially available product such as Microsoft's product .net. Additionally, rather
 than run every sequence through one and/or another formula, one may compare a
 subset of the sequences, which may be desirable if for example only a subset are
 available. For instance, it may be desirable to first perform a BLAST (Basic Local
Alignment Search Tool) search and to identify sequences that have no homology to
 25 other targets. Alternatively, it may be desirable to scan the sequence and to identify
 regions of moderate GC context, then perform relevant calculations using one of the
 above-described formulas on these regions. These calculations can be done manually
 or with the aid of a computer.

30 As with Formulas I – VII, either Formula VIII or Formula IX may be used for
 a given mRNA target sequence. However, it is possible that according to one or the
 other formula more than one siRNA will have the same value. Accordingly, it is
 beneficial to have a second formula by which to differentiate sequences. Formula IX
 was derived in a similar fashion as Formula VIII, yet used a larger data set and thus

yields sequences with higher statistical correlations to highly functional duplexes. The sequence that has the highest value ascribed to it may be referred to as a "first optimized duplex." The sequence that has the second highest value ascribed to it may be referred to as a "second optimized duplex." Similarly, the sequences that have the third and fourth highest values ascribed to them may be referred to as a third optimized duplex and a fourth optimized duplex, respectively. When more than one sequence has the same value, each of them may, for example, be referred to as first optimized duplex sequences or co-first optimized duplexes.

10 SiRNA sequences identified using Formula VIII are contained within the enclosed compact disks. The data included on the enclosed compact disks is described more fully below. The sequences identified by Formula VIII that are disclosed in the compacts disks may be used in gene silencing applications.

15 It should be noted that for Formulas VIII and IX all of the aforementioned criteria are identified as positions on the sense strand when oriented in the 5' to 3' direction as they are identified in connection with Formulas I – VII unless otherwise specified.

20 Formulas I - IX, may be used to select or to evaluate one, or more than one, siRNA in order to optimize silencing. Preferably, at least two optimized siRNAs that have been selected according to at least one of these formulas are used to silence a gene, more preferably at least three and most preferably at least four. The siRNAs may be used individually or together in a pool or kit. Further, they may be applied to a cell simultaneously or separately. Preferably, the at least two siRNAs are applied simultaneously. Pools are particularly beneficial for many research applications. However, for therapeutics, it may be more desirable to employ a single hyperfunctional siRNA as described elsewhere in this application.

25 When planning to conduct gene silencing, and it is necessary to choose between two or more siRNAs, one should do so by comparing the relative values when the siRNA are subjected to one of the formulas above. In general a higher scored siRNA should be used.

Useful applications include, but are not limited to, target validation, gene functional analysis, research and drug discovery, gene therapy and therapeutics. Methods for using siRNA in these applications are well known to persons of skill in the art.

5

Because the ability of siRNA to function is dependent on the sequence of the RNA and not the species into which it is introduced, the present invention is applicable across a broad range of species, including but not limited to all mammalian species, such as humans, dogs, horses, cats, cows, mice, hamsters, chimpanzees and gorillas, as well as other species and organisms such as bacteria, viruses, insects, plants and *C. elegans*.

The present invention is also applicable for use for silencing a broad range of genes, including but not limited to the roughly 45,000 genes of a human genome, and has particular relevance in cases where those genes are associated with diseases such as diabetes, Alzheimer's, cancer, as well as all genes in the genomes of the aforementioned organisms.

The siRNA selected according to the aforementioned criteria or one of the aforementioned algorithms are also, for example, useful in the simultaneous screening and functional analysis of multiple genes and gene families using high throughput strategies, as well as in direct gene suppression or silencing.

Development of the Algorithms

To identify siRNA sequence features that promote functionality and to quantify the importance of certain currently accepted conventional factors—such as G/C content and target site accessibility—the inventors synthesized an siRNA panel consisting of 270 siRNAs targeting three genes, Human Cyclophilin, Firefly Luciferase, and Human DBI. In all three cases, siRNAs were directed against specific regions of each gene. For Human Cyclophilin and Firefly Luciferase, ninety siRNAs were directed against a 199 bp segment of each respective mRNA. For DBI, 90 siRNAs were directed against a smaller, 109 base pair region of the mRNA. The sequences to which the siRNAs were directed are provided below.

It should be noted that in certain sequences, “t” is present. This is because many databases contain information in this manner. However, the t denotes a uracil residue in mRNA and siRNA. Any algorithm will, unless otherwise specified, process a t in a sequence as a u.

5

Human cyclophilin: 193—390, M60857

SEQ. ID NO. 29:

gttccaaaacagtggataatttgtggccttagctacaggagagaaaggatttggctacaaaacagcaaattccatcgtgt
aatcaaggacttcatgatccagggcggagacttcaccaggggagatggcacaggaggaaagagcatctacgggtgagcg
10 cttccccgatgagaacttcaaactgaagcactacgggcctggctggg

Firefly luciferase: 1434—1631, U47298 (pGL3, Promega)

SEQ. ID NO. 30:

tgaacttcccgcgccgttgtgttttgagcacggaaagacgatgacggaaaaagagatcgtggattacgtcgccagtca
15 agtaacaaccgcgaaaaagttgcgcggaggagttgtgttttgagcgaagtaccgaaaggcttaccggaaaactcgacg
caagaaaaatcagagagatcctcataaaggccaagaagg

DBI, NM_020548 (202-310) (every position)

SEQ. ID NO. 0031:

20 acgggcaaggccaagtgggatgcctggaatgagctgaaagggacttccaaggaagatgccatgaaagcttacatcaaca
aagtagaagagctaaagaaaaatacggg

A list of the siRNAs appears in Table III (see Examples Section, Example II)

25

The set of duplexes was analyzed to identify correlations between siRNA functionality and other biophysical or thermodynamic properties. When the siRNA panel was analyzed in functional and non-functional subgroups, certain nucleotides were much more abundant at certain positions in functional or non-functional groups. More specifically, the frequency of each nucleotide at each position in highly
30 functional siRNA duplexes was compared with that of nonfunctional duplexes in order to assess the preference for or against any given nucleotide at every position. These analyses were used to determine important criteria to be included in the siRNA algorithms (Formulas VIII and IX).

The data set was also analyzed for distinguishing biophysical properties of siRNAs in the functional group, such as optimal percent of GC content, propensity for internal structures and regional thermodynamic stability. Of the presented criteria, several are involved in duplex recognition, RISC activation/duplex unwinding, and target cleavage catalysis.

The original data set that was the source of the statistically derived criteria is shown in **Figure 2**. Additionally, this figure shows that random selection yields siRNA duplexes with unpredictable and widely varying silencing potencies as measured in tissue culture using HEK293 cells. In the figure, duplexes are plotted such that each x-axis tick-mark represents an individual siRNA, with each subsequent siRNA differing in target position by two nucleotides for Human Cyclophilin and Firefly Luciferase, and by one nucleotide for Human DBI. Furthermore, the y-axis denotes the level of target expression remaining after transfection of the duplex into cells and subsequent silencing of the target.

SiRNA identified and optimized in this document work equally well in a wide range of cell types. **Figure 3a** shows the evaluation of thirty siRNAs targeting the DBI gene in three cell lines derived from different tissues. Each DBI siRNA displays very similar functionality in HEK293 (ATCC, CRE-1573, human embryonic kidney), HeLa (ATCC, CCL-2, cervical epithelial adenocarcinoma) and DU145 (HTB-81, prostate) cells as determined by the B-DNA assay. Thus, siRNA functionality is determined by the primary sequence of the siRNA and not by the intracellular environment. Additionally, it should be noted that although the present invention provides for a determination of the functionality of siRNA for a given target, the same siRNA may silence more than one gene. For example, the complementary sequence of the silencing siRNA may be present in more than one gene. Accordingly, in these circumstances, it may be desirable not to use the siRNA with highest SMARTscore™. In such circumstances, it may be desirable to use the siRNA with the next highest SMARTscore™.

To determine the relevance of G/C content in siRNA function, the G/C content of each duplex in the panel was calculated and the functional classes of siRNAs ($<F50$, $\geq F50$, $\geq F80$, $\geq F95$ where F refers to the percent gene silencing) were sorted

accordingly. The majority of the highly-functional siRNAs (\geq F95) fell within the G/C content range of 36%—52% (**Figure 3B**). Twice as many non-functional ($<$ F50) duplexes fell within the high G/C content groups ($>$ 57% GC content) compared to the 36%—52% group. The group with extremely low GC content (26% or less) contained a higher proportion of non-functional siRNAs and no highly-functional siRNAs. The G/C content range of 30%—52% was therefore selected as Criterion I for siRNA functionality, consistent with the observation that a G/C range 30%—70% promotes efficient RNAi targeting. Application of this criterion alone provided only a marginal increase in the probability of selecting functional siRNAs from the panel: selection of F50 and F95 siRNAs was improved by 3.6% and 2.2%, respectively. The siRNA panel presented here permitted a more systematic analysis and quantification of the importance of this criterion than that used previously.

A relative measure of local internal stability is the A/U base pair (bp) content; therefore, the frequency of A/U bp was determined for each of the five terminal positions of the duplex (5' sense (S)/5' antisense (AS)) of all siRNAs in the panel. Duplexes were then categorized by the number of A/U bp in positions 1—5 and 15—19 of the sense strand. The thermodynamic flexibility of the duplex 5'-end (positions 1—5; S) did not appear to correlate appreciably with silencing potency, while that of the 3'-end (positions 15—19; S) correlated with efficient silencing. No duplexes lacking A/U bp in positions 15—19 were functional. The presence of one A/U bp in this region conferred some degree of functionality, but the presence of three or more A/Us was preferable and therefore defined as Criterion II. When applied to the test panel, only a marginal increase in the probability of functional siRNA selection was achieved: a 1.8% and 2.3% increase for F50 and F95 duplexes, respectively (**Table IV**).

The complementary strands of siRNAs that contain internal repeats or palindromes may form internal fold-back structures. These hairpin-like structures exist in equilibrium with the duplexed form effectively reducing the concentration of functional duplexes. The propensity to form internal hairpins and their relative stability can be estimated by predicted melting temperatures. High T_m reflects a tendency to form hairpin structures. Lower T_m values indicate a lesser tendency to form hairpins. When the functional classes of siRNAs were sorted by T_m (**Figure 3c**),

30

the following trends were identified: duplexes lacking stable internal repeats were the most potent silencers (no F95 duplex with predicted hairpin structure $T_m > 60^\circ\text{C}$). In contrast, about 60% of the duplexes in the groups having internal hairpins with calculated T_m values less than 20°C were F80. Thus, the stability of internal repeats is inversely proportional to the silencing effect and defines Criterion III (predicted hairpin structure $T_m \leq 20^\circ\text{C}$).

Sequence-based determinants of siRNA functionality

When the siRNA panel was sorted into functional and non-functional groups, the frequency of a specific nucleotide at each position in a functional siRNA duplex was compared with that of a nonfunctional duplex in order to assess the preference for or against a certain nucleotide. **Figure 4** shows the results of these queries and the subsequent resorting of the data set (from **Figure 2**). The data is separated into two sets: those duplexes that meet the criteria, a specific nucleotide in a certain position - grouped on the left (Selected) and those that do not - grouped on the right (Eliminated). The duplexes are further sorted from most functional to least functional with the y-axis of **Figure 4a-e** representing the % expression *i.e.* the amount of silencing that is elicited by the duplex (Note: each position on the X-axis represents a different duplex). Statistical analysis revealed correlations between silencing and several sequence-related properties of siRNAs. **Figure 4** and **Table IV** show quantitative analysis for the following five sequence-related properties of siRNA: (A) an A at position 19 of the sense strand; (B) an A at position 3 of the sense strand; (C) a U at position 10 of the sense strand; (D) a base other than G at position 13 of the sense strand; and (E) a base other than C at position 19 of the sense strand.

When the siRNAs in the panel were evaluated for the presence of an A at position 19 of the sense strand, the percentage of non-functional duplexes decreased from 20% to 11.8%, and the percentage of F95 duplexes increased from 21.7% to 29.4% (**Table IV**). Thus, the presence of an A in this position defined Criterion IV.

Another sequence-related property correlated with silencing was the presence of an A in position 3 of the sense strand (**Figure 4b**). Of the siRNAs with A3, 34.4% were F95, compared with 21.7% randomly selected siRNAs. The presence of a U base in position 10 of the sense strand exhibited an even greater impact (**Figure 4c**).

Of the duplexes in this group, 41.7% were F95. These properties became criteria V and VI, respectively.

Two negative sequence-related criteria that were identified also appear on
5 **Figure 4**. The absence of a G at position 13 of the sense strand, conferred a marginal increase in selecting functional duplexes (**Figure 4d**). Similarly, lack of a C at position 19 of the sense strand also correlated with functionality (**Figure 4e**). Thus, among functional duplexes, position 19 was most likely occupied by A, and rarely occupied by C. These rules were defined as criteria VII and VIII, respectively.

10

Application of each criterion individually provided marginal but statistically significant increases in the probability of selecting a potent siRNA. Although the results were informative, the inventors sought to maximize potency and therefore consider multiple criteria or parameters. Optimization is particularly important when
15 developing therapeutics. Interestingly, the probability of selecting a functional siRNA based on each thermodynamic criteria was 2%—4% higher than random, but 4%—8% higher for the sequence-related determinates. Presumably, these sequence-related increases reflect the complexity of the RNAi mechanism and the multitude of protein-RNA interactions that are involved in RNAi-mediated silencing.

20

Table IV

	Criterion	% Functional		Improvement over Random
5	I. 30%—52% G/C content	< F50	16.4%	-3.6%
		≥ F50	83.6%	3.6%
		≥ F80	60.4%	4.3%
		≥ F95	23.9%	2.2%
10	II. At least 3 A/U bases at positions 15—19 of the sense strand	< F50	18.2%	-1.8%
		≥ F50	81.8%	1.8%
		≥ F80	59.7%	3.6%
		≥ F95	24.0%	2.3%
15	III. Absence of internal repeats, as measured by T_m of secondary structure $\leq 20^\circ\text{C}$	< F50	16.7%	-3.3%
		≥ F50	83.3%	3.3%
		≥ F80	61.1%	5.0%
		≥ F95	24.6%	2.9%
20	IV. An A base at position 19 of the sense strand	< F50	11.8%	-8.2%
		≥ F50	88.2%	8.2%
		≥ F80	75.0%	18.9%
		≥ F95	29.4%	7.7%
25	V. An A base at position 3 of the sense strand	< F50	17.2%	-2.8%
		≥ F50	82.8%	2.8%
		≥ F80	62.5%	6.4%
		≥ F95	34.4%	12.7%
30	VI. A U base at position 10 of the sense strand	< F50	13.9%	-6.1%
		≥ F50	86.1%	6.1%
		≥ F80	69.4%	13.3%
		≥ F95	41.7%	20%
	VII. A base other than C at position 19 of the sense strand	< F50	18.8%	-1.2%
		≥ F50	81.2%	1.2%
		≥ F80	59.7%	3.6%
		≥ F95	24.2%	2.5%
	VIII. A base other than G at position 13 of the sense strand	< F50	15.2%	-4.8%
		≥ F50	84.8%	4.8%
		≥ F80	61.4%	5.3%
		≥ F95	26.5%	4.8%

The siRNA selection algorithm

In an effort to improve selection further, all identified criteria, including but not limited to those listed in Table IV were combined into the algorithms embodied in Formula VIII and Formula IX. Each siRNA was then assigned a score (referred to as a SMARTscore™) according to the values derived from the formulas. Duplexes that scored higher than 0 or 20, for Formulas VIII and IX, respectively, effectively selected a set of functional siRNAs and excluded all non-functional siRNAs. Conversely, all duplexes scoring lower than 0 and 20 according to formulas VIII and

IX, respectively, contained some functional siRNAs but included all non-functional siRNAs. A graphical representation of this selection is shown in **Figure 5**.

The methods for obtaining the seven criteria embodied in Table IV are illustrative of the results of the process used to develop the information for Formulas VIII and IX. Thus similar techniques were used to establish the other variables and their multipliers. As described above, basic statistical methods were used to determine the relative values for these multipliers.

To determine the value for "Improvement over Random" the difference in the frequency of a given attribute (e.g. GC content, base preference) at a particular position is determined between individual functional groups (e.g. <F50) and the total siRNA population studied (e.g. 270 siRNA molecules selected randomly). Thus, for instance, in Criterion I (30%-52% GC content) members of the <F50 group were observed to have GC contents between 30-52% in 16.4% of the cases. In contrast, the total group of 270 siRNAs had GC contents in this range, 20% of the time. Thus for this particular attribute, there is a small negative correlation between 30%-52% GC content and this functional group (i.e. $16.4\% - 20\% = -3.6\%$). Similarly, for Criterion VI, (a "U" at position 10 of the sense strand), the >F95 group contained a "U" at this position 41.7% of the time. In contrast, the total group of 270 siRNAs had a "U" at this position 21.7% of the time, thus the improvement over random is calculated to be 20% (or $41.7\% - 21.7\%$).

Identifying The Average Internal Stability Profile of Strong siRNA

In order to identify an internal stability profile that is characteristic of strong siRNA, 270 different siRNAs derived from the cyclophilin B, the diazepam binding inhibitor (DBI), and the luciferase gene were individually transfected into HEK293 cells and tested for their ability to induce RNAi of the respective gene. Based on their performance in the *in vivo* assay, the sequences were then subdivided into three groups, (i) >95% silencing; (ii) 80-95% silencing; and (iii) less than 50% silencing. Sequences exhibiting 51-84% silencing were eliminated from further consideration to reduce the difficulties in identifying relevant thermodynamic patterns.

Following the division of siRNA into three groups, a statistical analysis was performed on each member of each group to determine the average internal stability profile (AISP) of the siRNA. To accomplish this the Oligo 5.0 Primer Analysis Software and other related statistical packages (*e.g.* Excel) were exploited to

5 determine the internal stability of pentamers using the nearest neighbor method described by Freier *et al.*, (1986) *Improved free-energy parameters for predictions of RNA duplex stability*, Proc Natl. Acad. Sci. U. S. A. 83(24): 9373-7. Values for each group at each position were then averaged, and the resulting data were graphed on a linear coordinate system with the Y-axis expressing the ΔG (free energy) values in

10 kcal/mole and the X-axis identifying the position of the base relative to the 5' end.

The results of the analysis identified multiple key regions in siRNA molecules that were critical for successful gene silencing. At the 3'-most end of the sense strand (5'antisense), highly functional siRNA (>95% gene silencing, see Figure 6a, >F95)

15 have a low internal stability (AISP of position 19 = ~ -7.6 kcal/mol). In contrast low-efficiency siRNA (*i.e.* those exhibiting less than 50% silencing, <F50) display a distinctly different profile, having high ΔG values (~ -8.4 kcal/mol) for the same position. Moving in a 5' (sense strand) direction, the internal stability of highly efficient siRNA rises (position 12 = ~ -8.3 kcal/mole) and then drops again (position 7

20 = ~ -7.7 kcal/mol) before leveling off at a value of approximately -8.1 kcal/mol for the 5' terminus. SiRNA with poor silencing capabilities show a distinctly different profile. While the AISP value at position 12 is nearly identical with that of strong siRNAs, the values at positions 7 and 8 rise considerably, peaking at a high of ~ -9.0 kcal/mol. In addition, at the 5' end of the molecule the AISP profile of strong and

25 weak siRNA differ dramatically. Unlike the relatively strong values exhibited by siRNA in the >95% silencing group, siRNAs that exhibit poor silencing activity have weak AISP values (-7.6 , -7.5 , and -7.5 kcal/mol for positions 1, 2 and 3 respectively).

Overall the profiles of both strong and weak siRNAs form distinct sinusoidal

30 shapes that are roughly 180° out-of-phase with each other. While these thermodynamic descriptions define the archetypal profile of a strong siRNA, it will likely be the case that neither the ΔG values given for key positions in the profile or the absolute position of the profile along the Y-axis (*i.e.* the ΔG -axis) are absolutes.

Profiles that are shifted upward or downward (*i.e.* having on an average, higher or lower values at every position) but retain the relative shape and position of the profile along the X-axis can be foreseen as being equally effective as the model profile described here. Moreover, it is likely that siRNA that have strong or even stronger gene-specific silencing effects might have exaggerated ΔG values (either higher or lower) at key positions. Thus, for instance, it is possible that the 5'-most position of the sense strand (position 19) could have ΔG values of 7.4 kcal/mol or lower and still be a strong siRNA if, for instance, a G-C \rightarrow G-T/U mismatch were substituted at position 19 and altered duplex stability. Similarly, position 12 and position 7 could have values above 8.3 kcal/mol and below 7.7 kcal/mole, respectively, without abating the silencing effectiveness of the molecule. Thus, for instance, at position 12, a stabilizing chemical modification (*e.g.* a chemical modification of the 2' position of the sugar backbone) could be added that increases the average internal stability at that position. Similarly, at position 7, mismatches similar to those described previously could be introduced that would lower the ΔG values at that position.

Lastly, it is important to note that while functional and non-functional siRNA were originally defined as those molecules having specific silencing properties, both broader or more limiting parameters can be used to define these molecules. As used herein, unless otherwise specified, "non-functional siRNA" are defined as those siRNA that induce less than 50% (<50%) target silencing, "semi-functional siRNA" induce 50-79% target silencing, "functional siRNA" are molecules that induce 80-95% gene silencing, and "highly-functional siRNA" are molecules that induce greater than 95% gene silencing. These definitions are not intended to be rigid and can vary depending upon the design and needs of the application. For instance, it is possible that a researcher attempting to map a gene to a chromosome using a functional assay, may identify an siRNA that reduces gene activity by only 30%. While this level of gene silencing may be "non-functional" for *e.g.* therapeutic needs, it is sufficient for gene mapping purposes and is, under these uses and conditions, "functional." For these reasons, functional siRNA can be defined as those molecules having greater than 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% silencing capabilities at 100nM transfection conditions. Similarly, depending upon the needs of the study and/or application, non-functional and semi-functional siRNA can be defined as

having different parameters. For instance, semi-functional siRNA can be defined as being those molecules that induce 20%, 30%, 40%, 50%, 60%, or 70% silencing at 100nM transfection conditions. Similarly, non-functional siRNA can be defined as being those molecules that silence gene expression by less than 70%, 60%, 50%,
5 40%, 30%, or less. Nonetheless, unless otherwise stated, the descriptions stated in the "Definitions" section of this text should be applied.

Functional attributes can be assigned to each of the key positions in the AISP of strong siRNA. The low 5' (sense strand) AISP values of strong siRNAs may be
10 necessary for determining which end of the molecule enters the RISC complex. In contrast, the high and low AISP values observed in the central regions of the molecule may be critical for siRNA-target mRNA interactions and product release, respectively.

15 If the AISP values described above accurately define the thermodynamic parameters of strong siRNA, it would be expected that similar patterns would be observed in strong siRNA isolated from nature. Natural siRNAs exist in a harsh, RNase-rich environment and it can be hypothesized that only those siRNA that exhibit heightened affinity for RISC (*i.e.* siRNA that exhibit an average internal
20 stability profile similar to those observed in strong siRNA) would survive in an intracellular environment. This hypothesis was tested using GFP-specific siRNA isolated from *N. benthamiana*. Llave *et al.* (2002) *Endogenous and Silencing-Associated Small RNAs in Plants*, *The Plant Cell* 14, 1605-1619, introduced long double-stranded GFP-encoding RNA into plants and subsequently re-isolated GFP-
25 specific siRNA from the tissues. The AISP of fifty-nine of these GFP-siRNA were determined, averaged, and subsequently plotted alongside the AISP profile obtained from the cyclophilin B/DBI/ luciferase siRNA having >90% silencing properties (**Figure 6b**). Comparison of the two groups show that profiles are nearly identical. This finding validates the information provided by the internal stability profiles and
30 demonstrates that: (1) the profile identified by analysis of the cyclophilin B/DBI/ luciferase siRNAs are not gene specific; and (2) AISP values can be used to search for strong siRNAs in a variety of species.

Both chemical modifications and base-pair mismatches can be incorporated into siRNA to alter the duplex's AISP and functionality. For instance, introduction of mismatches at positions 1 or 2 of the sense strand destabilized the 5' end of the sense strand and increases the functionality of the molecule (see Luc, **Figure 7**). Similarly, addition of 2'-O-methyl groups to positions 1 and 2 of the sense strand can also alter the AISP and (as a result) increase both the functionality of the molecule and eliminate off-target effects that results from sense strand homology with the unrelated targets (**Figures 8a, 8b**).

10 Rationale for Criteria in a Biological Context

The fate of siRNA in the RNAi pathway may be described in 5 major steps: (1) duplex recognition and pre-RISC complex formation; (2) ATP-dependent duplex unwinding/strand selection and RISC activation; (3) mRNA target identification; (4) mRNA cleavage, and (5) product release (**Figure 1**). Given the level of nucleic acid-protein interactions at each step, siRNA functionality is likely influenced by specific biophysical and molecular properties that promote efficient interactions within the context of the multi-component complexes. Indeed, the systematic analysis of the siRNA test set identified multiple factors that correlate well with functionality. When combined into a single algorithm, they proved to be very effective in selecting active siRNAs.

The factors described here may also be predictive of key functional associations important for each step in RNAi. For example, the potential formation of internal hairpin structures correlated negatively with siRNA functionality.

25 Complementary strands with stable internal repeats are more likely to exist as stable hairpins thus decreasing the effective concentration of the functional duplex form. This suggests that the duplex is the preferred conformation for initial pre-RISC association. Indeed, although single complementary strands can induce gene silencing, the effective concentration required is at least two orders of magnitude

30 higher than that of the duplex form.

siRNA-pre-RISC complex formation is followed by an ATP-dependent duplex unwinding step and "activation" of the RISC. The siRNA functionality was shown to correlate with overall low internal stability of the duplex and low internal stability of

the 3' sense end (or differential internal stability of the 3' sense compare to the 5' sense strand), which may reflect strand selection and entry into the RISC. Overall duplex stability and low internal stability at the 3' end of the sense strand were also correlated with siRNA functionality. Interestingly, siRNAs with very high and very low overall stability profiles correlate strongly with non-functional duplexes. One interpretation is that high internal stability prevents efficient unwinding while very low stability reduces siRNA target affinity and subsequent mRNA cleavage by the RISC.

Several criteria describe base preferences at specific positions of the sense strand and are even more intriguing when considering their potential mechanistic roles in target recognition and mRNA cleavage. Base preferences for A at position 19 of the sense strand but not C, are particularly interesting because they reflect the same base preferences observed for naturally occurring miRNA precursors. That is, among the reported miRNA precursor sequences 75% contain a U at position 1 which corresponds to an A in position 19 of the sense strand of siRNAs, while G was under-represented in this same position for miRNA precursors. These observations support the hypothesis that both miRNA precursors and siRNA duplexes are processed by very similar if not identical protein machinery. The functional interpretation of the predominance of a U/A base pair is that it promotes flexibility at the 5' antisense ends of both siRNA duplexes and miRNA precursors and facilitates efficient unwinding and selective strand entrance into an activated RISC.

Among the criteria associated with base preferences that are likely to influence mRNA cleavage or possibly product release, the preference for U at position 10 of the sense strand exhibited the greatest impact, enhancing the probability of selecting an F80 sequence by 13.3%. Activated RISC preferentially cleaves target mRNA between nucleotides 10 and 11 relative to the 5' end of the complementary targeting strand. Therefore, it may be that U, the preferred base for most endoribonucleases, at this position supports more efficient cleavage. Alternatively, a U/A bp between the targeting siRNA strand and its cognate target mRNA may create an optimal conformation for the RISC-associated "slicing" activity.

Pooling

According to another embodiment, the present invention provides a pool of at least two siRNAs, preferably in the form of a kit or therapeutic reagent, wherein one strand of each of the siRNAs, the sense strand comprises a sequence that is

5 substantially similar to a sequence within a target mRNA. The opposite strand, the antisense strand, will preferably comprise a sequence that is substantially complementary to that of the target mRNA. More preferably, one strand of each siRNA will comprise a sequence that is identical to a sequence that is contained in the target mRNA. Most preferably, each siRNA will be 19 base pairs in length, and one

10 strand of each of the siRNAs will be 100% complementary to a portion of the target mRNA.

By increasing the number of siRNAs directed to a particular target using a pool or kit, one is able both to increase the likelihood that at least one siRNA with

15 satisfactory functionality will be included, as well as to benefit from additive or synergistic effects. Further, when two or more siRNAs directed against a single gene do not have satisfactory levels of functionality alone, if combined, they may satisfactorily promote degradation of the target messenger RNA and successfully inhibit translation. By including multiple siRNAs in the system, not only is the

20 probability of silencing increased, but the economics of operation are also improved when compared to adding different siRNAs sequentially. This effect is contrary to the conventional wisdom that the concurrent use of multiple siRNA will negatively impact gene silencing (*e.g.* Holen, T. *et al.* (2003) "Similar behavior of single strand and double strand siRNAs suggests they act through a common RNAi pathway."

25 NAR 31: 2401-21407).

In fact, when two siRNAs were pooled together, 54% of the pools of two siRNAs induced more than 95% gene silencing. Thus, a 2.5-fold increase in the percentage of functionality was achieved by randomly combining two siRNAs.

30 Further, over 84% of pools containing two siRNAs induced more than 80% gene silencing.

More preferably, the kit is comprised of at least three siRNAs, wherein one strand of each siRNA comprises a sequence that is substantially similar to a sequence

of the target mRNA and the other strand comprises a sequence that is substantially complementary to the region of the target mRNA. As with the kit that comprises at least two siRNAs, more preferably one strand will comprise a sequence that is identical to a sequence that is contained in the mRNA and another strand that is 100% complementary to a sequence that is contained in the mRNA. During experiments, when three siRNAs were combined together, 60% of the pools induced more than 95% gene silencing and 92% of the pools induced more than 80% gene silencing.

Further, even more preferably, the kit is comprised of at least four siRNAs, wherein one strand of each siRNA comprises a sequence that is substantially similar to a region of the sequence of the target mRNA, and the other strand comprises a sequence that is substantially complementary to the region of the target mRNA. As with the kit or pool that comprises at least two siRNAs, more preferably one strand of each of the siRNA duplexes will comprise a sequence that is identical to a sequence that is contained in the mRNA, and another strand that is 100% complementary to a sequence that is contained in the mRNA.

Additionally, kits and pools with at least five, at least six, and at least seven siRNAs may also be useful with the present invention. For example, pools of five siRNA induced 95% gene silencing with 77% probability and 80% silencing with 98.8% probability. Thus, pooling of siRNAs together can result in the creation of a target-specific silencing reagent with almost a 99% probability of being functional. The fact that such high levels of success are achievable using such pools of siRNA, enables one to dispense with costly and time-consuming target-specific validation procedures.

For this embodiment, as well as the other aforementioned embodiments, each of the siRNAs within a pool will preferably comprise between 18 and 30 base pairs, more preferably between 18 and 25 base pairs, and most preferably 19 base pairs. Within each siRNA, preferably at least 18 contiguous bases of the antisense strand will be 100% complementary to the target mRNA. More preferably, at least 19 contiguous bases of the antisense strand will be 100% complementary to the target mRNA. Additionally, there may be overhangs on either the sense strand or the antisense strand, and these overhangs may be at either the 5' end or the 3' end of

either of the strands, for example there may be one or more overhangs of 1-6 bases. When overhangs are present, they are not included in the calculation of the number of base pairs. The two nucleotide 3' overhangs mimic natural siRNAs and are commonly used but are not essential. Preferably, the overhangs should consist of two nucleotides, most often dTdT or UU at the 3' end of the sense and antisense strand that are not complementary to the target sequence. The siRNAs may be produced by any method that is now known or that comes to be known for synthesizing double stranded RNA that one skilled in the art would appreciate would be useful in the present invention. Preferably, the siRNAs will be produced by Dharmacon's proprietary ACE® technology. However, other methods for synthesizing siRNAs are well known to persons skilled in the art and include, but are not limited to, any chemical synthesis of RNA oligonucleotides, ligation of shorter oligonucleotides, *in vitro* transcription of RNA oligonucleotides, the use of vectors for expression within cells, recombinant Dicer products and PCR products.

The siRNA duplexes within the aforementioned pools of siRNAs may correspond to overlapping sequences within a particular mRNA, or non-overlapping sequences of the mRNA. However, preferably they correspond to non-overlapping sequences. Further, each siRNA may be selected randomly, or one or more of the siRNA may be selected according to the criteria discussed above for maximizing the effectiveness of siRNA.

Included in the definition of siRNAs are siRNAs that contain substituted and/or labeled nucleotides that may, for example, be labeled by radioactivity, fluorescence or mass. The most common substitutions are at the 2' position of the ribose sugar, where moieties such as H (hydrogen) F, NH₃, OCH₃ and other O- alkyl, alkenyl, alkynyl, and orthoesters, may be substituted, or in the phosphorous backbone, where sulfur, amines or hydrocarbons may be substituted for the bridging of non-bridging atoms in the phosphodiester bond. Examples of modified siRNAs are explained more fully in commonly assigned U.S. Patent Application Ser. No. 10/613,077, filed July 1, 2003, which is incorporated by reference herein.

Additionally, as noted above, the cell type into which the siRNA is introduced may affect the ability of the siRNA to enter the cell; however, it does not appear to

affect the ability of the siRNA to function once it enters the cell. Methods for introducing double-stranded RNA into various cell types are well known to persons skilled in the art.

5 As persons skilled in the art are aware, in certain species, the presence of proteins such as RdRP, the RNA-dependent RNA polymerase, may catalytically enhance the activity of the siRNA. For example, RdRP propagates the RNAi effect in *C. elegans* and other non-mammalian organisms. In fact, in organisms that contain these proteins, the siRNA may be inherited. Two other proteins that are well studied
10 and known to be a part of the machinery are members of the Argonaute family and Dicer, as well as their homologues. There is also initial evidence that the RISC complex might be associated with the ribosome so the more efficiently translated mRNAs will be more susceptible to silencing than others.

15 Another very important factor in the efficacy of siRNA is mRNA localization. In general, only cytoplasmic mRNAs are considered to be accessible to RNAi to any appreciable degree. However, appropriately designed siRNAs, for example, siRNAs modified with internucleotide linkages, may be able to cause silencing by acting in the nucleus. Examples of these types of modifications are described in commonly
20 assigned U.S. Patent Application Serial Nos. 10/431,027 and 10/613,077, each of which is incorporated by reference herein.

 As described above, even when one selects ~~at least two~~ siRNAs at random, the effectiveness of the two may be greater than one would predict based on the
25 effectiveness of two individual siRNAs. This additive or synergistic effect is particularly noticeable as one increases to at least three siRNAs, and even more noticeable as one moves to at least four siRNAs. Surprisingly, the pooling of the non-functional and semi-functional siRNAs, particularly more than five siRNAs, can lead to a silencing mixture that is as effective if not more effective than any one particular
30 functional siRNA.

Within the kit of the present invention, preferably each siRNA will be present in a concentration of between 0.001 and 200 μ M, more preferably between 0.01 and 200 nM, and most preferably between 0.1 and 10 nM.

5 In addition to preferably comprising at least four or five siRNAs, the kit of the present invention will also preferably comprise a buffer to keep the siRNA duplex stable. Persons skilled in the art are aware of buffers suitable for keeping siRNA stable. For example, the buffer may be comprised of 100 mM KCl, 30 mM HEPES-
10 pH 7.5, and 1 mM $MgCl_2$. Alternatively, kits might contain complementary strands that contain any one of a number of chemical modifications (e.g. a 2'-O-ACE) that protect the agents from degradation by nucleases. In this instance, the user may (or may not) remove the modifying protective group (e.g. deprotect) before annealing the two complementary strands together.

15 By way of example, the kit may be organized such that pools of siRNA duplexes are provided on an array or microarray of wells or drops for a particular gene set or for unrelated genes. The array may, for example, be in 96 wells, 384 wells or 1284 wells arrayed in a plastic plate or on a glass slide using techniques now known or that come to be known to persons skilled in the art. Within an array,
20 preferably there will be controls such as functional anti-lamin A/C, cyclophilin and two siRNA duplexes that are not specific to the gene of interest.

 In order to ensure stability of the siRNA pools prior to usage, they may be retained in lyophilized form at minus twenty degrees ($-20^{\circ}C$) until they are ready for
25 use. Prior to usage, they should be resuspended; however, even once resuspended, for example, in the aforementioned buffer, they should be kept at minus twenty degrees, ($-20^{\circ}C$) until used. The aforementioned buffer, prior to use, may be stored at approximately $4^{\circ}C$ or room temperature. Effective temperatures at which to conduct transfections are well known to persons skilled in the art and include for example,
30 room temperature.

 The kit may be applied either *in vivo* or *in vitro*. Preferably, the siRNA of the pools or kits is applied to a cell through transfection, employing standard transfection

protocols. These methods are well known to persons skilled in the art and include the use of lipid-based carriers, electroporation, cationic carriers, and microinjection. Further, one could apply the present invention by synthesizing equivalent DNA sequences (either as two separate, complementary strands, or as hairpin molecules) instead of siRNA sequences and introducing them into cells through vectors. Once in the cells, the cloned DNA could be transcribed, thereby forcing the cells to generate the siRNA. Examples of vectors suitable for use with the present application include but are not limited to the standard transient expression vectors, adenoviruses, retroviruses, lentivirus-based vectors, as well as other traditional expression vectors. Any vector that has an adequate siRNA expression and procession module may be used. Furthermore, certain chemical modifications to siRNAs, including but not limited to conjugations to other molecules, may be used to facilitate delivery. For certain applications it may be preferable to deliver molecules without transfection by simply formulating in a physiological acceptable solution.

15

This embodiment may be used in connection with any of the aforementioned embodiments. Accordingly, the sequences within any pool may be selected by rational design.

20

Multigene Silencing

In addition to developing kits that contain multiple siRNA directed against a single gene, another embodiment includes the use of multiple siRNA targeting multiple genes. Multiple genes may be targeted through the use of high- or hyper-functional siRNA. High- or hyper- functional siRNA that exhibit increased potency, require lower concentrations to induce desired phenotypic (and thus therapeutic) effects. This circumvents RISC saturation. It therefore reasons that if lower concentrations of a single siRNA are needed for knockout or knockdown expression of one gene, then the remaining (uncomplexed) RISC will be free and available to interact with siRNA directed against two, three, four, or more, genes. Thus in this embodiment, the authors describe the use of highly functional or hyper-functional siRNA to knock out three separate genes. More preferably, such reagents could be combined to knockout four distinct genes. Even more preferably, highly functional or hyperfunctional siRNA could be used to knock out five distinct genes. Most

30

preferably, siRNA of this type could be used to knockout or knockdown the expression of six or more genes.

Hyperfunctional siRNA

- 5 The term hyperfunctional siRNA (hf-siRNA) describes a subset of the siRNA population that induces RNAi in cells at low- or sub-nanomolar concentrations for extended periods of time. These traits, heightened potency and extended longevity of the RNAi phenotype, are highly attractive from a therapeutic standpoint. Agents having higher potency require lesser amounts of the molecule to achieve the desired
- 10 physiological response, thus reducing the probability of side effects due to “off-target” interference. In addition to the potential therapeutic benefits associated with hyperfunctional siRNA, hf-siRNA are also desirable from an economic position. Hyperfunctional siRNA may cost less on a per-treatment basis, thus reducing the overall expenditures to both the manufacturer and the consumer.
- 15 Identification of hyperfunctional siRNA involves multiple steps that are designed to examine an individual siRNA agent’s concentration- and/or longevity-profiles. In one non-limiting example, a population of siRNA directed against a single gene are first analyzed using the previously described algorithm (Formula VIII). Individual siRNA are then introduced into a test cell line and assessed for the ability
- 20 to degrade the target mRNA. It is important to note that when performing this step it is not necessary to test all of the siRNA. Instead, it is sufficient to test only those siRNA having the highest SMARTscores™ (*i.e.* SMARTscore™ > -10). Subsequently, the gene silencing data is plotted against the SMARTscores™ (see Figure 9). SiRNA that (1) induce a high degree of gene silencing (*i.e.* they induce
- 25 greater than 80% gene knockdown) and (2) have superior SMARTscores™ (*i.e.* a SMARTscore™ of > -10, suggesting a desirable average internal stability profile) are selected for further investigations designed to better understand the molecule’s potency and longevity. In one, non-limiting study dedicated to understanding a molecule’s potency, an siRNA is introduced into one (or more) cell types in
- 30 increasingly diminishing concentrations (*e.g.* 3.0 → 0.3 nM). Subsequently, the level of gene silencing induced by each concentration is examined and siRNA that exhibit hyperfunctional potency (*i.e.* those that induce 80% silencing or greater at *e.g.* picomolar concentrations) are identified. In a second study, the longevity profiles of

siRNA having high (>-10) SMARTscores™ and greater than 80% silencing are examined. In one non-limiting example of how this is achieved, siRNA are introduced into a test cell line and the levels of RNAi are measured over an extended period of time (*e.g.* 24-168 hrs). SiRNAs that exhibit strong RNA interference patterns (*i.e.* $>80\%$ interference) for periods of time greater than, *e.g.*, 120 hours are thus identified. Studies similar to those described above can be performed on any and all of the $>10^6$ siRNA included in this document to further define the most functional molecule for any given gene. Molecules possessing one or both properties (extended longevity and heightened potency) are labeled “hyperfunctional siRNA,” and
10 earmarked as candidates for future therapeutic studies.

While the example(s) given above describe one means by which hyperfunctional siRNA can be isolated, neither the assays themselves nor the selection parameters used are rigid and can vary with each family of siRNA. Families
15 of siRNA include siRNAs directed against a single gene, or directed against a related family of genes.

The highest quality siRNA achievable for any given gene may vary considerably. Thus, for example, in the case of one gene (gene X), rigorous studies
20 such as those described above may enable the identification of an siRNA that, at picomolar concentrations, induces 99% silencing for a period of 10 days. Yet identical studies of a second gene (gene Y) may yield an siRNA that at high nanomolar concentrations (*e.g.* 100nM) induces only 75% silencing for a period of 2 days. Both molecules represent the very optimum siRNA for their respective gene
25 targets and therefore are designated “hyperfunctional.” Yet due to a variety of factors including but not limited to target concentration, siRNA stability, cell type, off-target interference, and others, equivalent levels of potency and longevity are not achievable. Thus, for these reasons, the parameters described in the before mentioned assays, can vary. While the initial screen selected siRNA that had SMARTscores™
30 above -10 and a gene silencing capability of greater than 80%, selections that have stronger (or weaker) parameters can be implemented. Similarly, in the subsequent studies designed to identify molecules with high potency and longevity, the desired cutoff criteria (*i.e.* the lowest concentration that induces a desirable level of interference, or the longest period of time that interference can be observed) can vary.

The experimentation subsequent to application of the rational criteria of this application is significantly reduced where one is trying to obtain a suitable hyperfunctional siRNA for, for example, therapeutic use. When, for example, the additional experimentation of the type described herein is applied by one skilled in the art with this disclosure in hand, a hyperfunctional siRNA is readily identified.

The siRNA may be introduced into a cell by any method that is now known or that comes to be known and that from reading this disclosure, persons skilled in the art would determine would be useful in connection with the present invention in enabling siRNA to cross the cellular membrane. These methods include, but are not limited to, any manner of transfection, such as for example transfection employing DEAE-Dextran, calcium phosphate, cationic lipids/liposomes, micelles, manipulation of pressure, microinjection, electroporation, immunoporation, use of vectors such as viruses, plasmids, cosmids, bacteriophages, cell fusions, and coupling of the polynucleotides to specific conjugates or ligands such as antibodies, antigens, or receptors, passive introduction, adding moieties to the siRNA that facilitate its uptake, and the like.

Having described the invention with a degree of particularity, examples will now be provided. These examples are not intended to and should not be construed to limit the scope of the claims in any way.

Examples

25 **General Techniques and Nomenclatures**

siRNA nomenclature. All siRNA duplexes are referred to by sense strand. The first nucleotide of the 5'-end of the sense strand is position 1, which corresponds to position 19 of the antisense strand for a 19-mer. In most cases, to compare results from different experiments, silencing was determined by measuring specific transcript mRNA levels or enzymatic activity associated with specific transcript levels, 24 hours post-transfection, with siRNA concentrations held constant at 100 nM. For all experiments, unless otherwise specified transfection efficiency was ensured to be over 95%, and no detectable cellular toxicity was observed. The following system of nomenclature was used to compare and report siRNA-silencing functionality: "F"

followed by the degree of minimal knockdown. For example, F50 signifies at least 50% knockdown, F80 means at least 80%, and so forth. For this study, all sub-F50 siRNAs were considered non-functional.

- 5 Cell culture and transfection. 96-well plates are coated with 50 μ L of 50 mg/ml poly-L-lysine (Sigma) for 1 hr, and then washed 3X with distilled water before being dried for 20 min. HEK293 cells or HEK293Lucs or any other cell type of interest are released from their solid support by trypsinization, diluted to 3.5×10^5 cells/ml, followed by the addition of 100 μ L of cells/well. Plates are then incubated overnight
10 at 37° C, 5% CO₂. Transfection procedures can vary widely depending on the cell type and transfection reagents. In one non-limiting example, a transfection mixture consisting of 2 mL Opti-MEM I (Gibco-BRL), 80 μ L Lipofectamine 2000 (Invitrogen), 15 μ L SUPERNasin at 20 U/ μ L (Ambion), and 1.5 μ L of reporter gene plasmid at 1 μ g/ μ L is prepared in 5-ml polystyrene round bottom tubes. 100 μ L of
15 transfection reagent is then combined with 100 μ L of siRNAs in polystyrene deep-well titer plates (Beckman) and incubated for 20 to 30 min at room temp. 550 μ L of Opti-MEM is then added to each well to bring the final siRNA concentration to 100 nM. Plates are then sealed with parafilm and mixed. Media is removed from HEK293 cells and replaced with 95 μ L of transfection mixture. Cells are incubated overnight at
20 37° C, 5% CO₂.

- Quantification of gene knockdown. A variety of quantification procedures can be used to measure the level of silencing induced by siRNA or siRNA pools. In one non-limiting example: to measure mRNA levels 24 hrs post-transfection, QuantiGene
25 branched-DNA (bDNA) kits (Bayer) (Wang, *et al*, *Regulation of insulin preRNA splicing by glucose*. Proc Natl Acad Sci 1997, 94:4360.) are used according to manufacturer instructions. To measure luciferase activity, media is removed from HEK293 cells 24 hrs post-transfection, and 50 μ L of Steady-GLO reagent (Promega) is added. After 5 min, plates are analyzed on a plate reader.

30

Example I. Sequences Used to Develop the Algorithm.

Anti-Firefly and anti-Cyclophilin siRNAs panels (Figure 5a, b) sorted according to using Formula VIII predicted values. All siRNAs scoring more than 0

(formula VIII) and more than 20 (formula IX) are fully functional. All ninety sequences for each gene (and DBI) appear below in **Table III**.

5

TABLE III

Cyclo	1	SEQ. ID 0032	GUUCCAAAAACAGUGGAUA
Cyclo	2	SEQ. ID 0033	UCCAAAAACAGUGGAUAAU
Cyclo	3	SEQ. ID 0034	CAAAAACAGUGGAUAAUUU
Cyclo	4	SEQ. ID 0035	AAAACAGUGGAUAAUUUUG
Cyclo	5	SEQ. ID 0036	AACAGUGGAUAAUUUUGUG
Cyclo	6	SEQ. ID 0037	CAGUGGAUAAUUUUGUGGC
Cyclo	7	SEQ. ID 0038	GUGGAUAAUUUUGUGGCCU
Cyclo	8	SEQ. ID 0039	GGAUAAUUUUGUGGCCUUA
Cyclo	9	SEQ. ID 0040	AUAAUUUUGUGGCCUAGC
Cyclo	10	SEQ. ID 0041	AAUUUUGUGGCCUAGCUA
Cyclo	11	SEQ. ID 0042	UUUUGUGGCCUAGCUACA
Cyclo	12	SEQ. ID 0043	UUGUGGCCUAGCUACAGG
Cyclo	13	SEQ. ID 0044	GUGGCCUAGCUACAGGAG
Cyclo	14	SEQ. ID 0045	GGCCUAGCUACAGGAGAG
Cyclo	15	SEQ. ID 0046	CCUAGCUACAGGAGAGAA
Cyclo	16	SEQ. ID 0047	UUAGCUACAGGAGAGAAAG
Cyclo	17	SEQ. ID 0048	AGCUACAGGAGAGAAAGGA
Cyclo	18	SEQ. ID 0049	CUACAGGAGAGAAAGGAUU
Cyclo	19	SEQ. ID 0050	ACAGGAGAGAAAGGAUUUG
Cyclo	20	SEQ. ID 0051	AGGAGAGAAAGGAUUUGGC
Cyclo	21	SEQ. ID 0052	GAGAGAAAGGAUUUGGCUA
Cyclo	22	SEQ. ID 0053	GAGAAAGGAUUUGGCUACA
Cyclo	23	SEQ. ID 0054	GAAAGGAUUUGGCUACAAA
Cyclo	24	SEQ. ID 0055	AAGGAUUUGGCUACAAAAA
Cyclo	25	SEQ. ID 0056	GGAUUUGGCUACAAAAACA
Cyclo	26	SEQ. ID 0057	AUUUGGCUACAAAAACAGC
Cyclo	27	SEQ. ID 0058	UUGGCUACAAAAACAGCAA
Cyclo	28	SEQ. ID 0059	GGCUACAAAAACAGCAAUU
Cyclo	29	SEQ. ID 0060	CUACAAAAACAGCAAUUUC
Cyclo	30	SEQ. ID 0061	ACAAAAACAGCAAUUUCCA
Cyclo	31	SEQ. ID 0062	AAAAACAGCAAUUUCCAUC
Cyclo	32	SEQ. ID 0063	AAACAGCAAUUUCCAUCGU
Cyclo	33	SEQ. ID 0064	ACAGCAAUUUCCAUCGUGU
Cyclo	34	SEQ. ID 0065	AGCAAUUUCCAUCGUGUAA

Cyclo	35	SEQ. ID 0066	CAAAUCCAUCGUGUAAUC
Cyclo	36	SEQ. ID 0067	AAUCCAUCGUGUAAUCAA
Cyclo	37	SEQ. ID 0068	UCCAUCGUGUAAUCAAGG
Cyclo	38	SEQ. ID 0069	CCAUCGUGUAAUCAAGGAC
Cyclo	39	SEQ. ID 0070	AUCGUGUAAUCAAGGACUU
Cyclo	40	SEQ. ID 0071	CGUGUAAUCAAGGACUUCA
Cyclo	41	SEQ. ID 0072	UGUAAUCAAGGACUUCAUG
Cyclo	42	SEQ. ID 0073	UAAUCAAGGACUUCAUGAU
Cyclo	43	SEQ. ID 0074	AUCAAGGACUUCAUGAUCC
Cyclo	44	SEQ. ID 0075	CAAGGACUUCAUGAUCCAG
Cyclo	45	SEQ. ID 0076	AGGACUUCAUGAUCCAGGG
Cyclo	46	SEQ. ID 0077	GACUUCAUGAUCCAGGGCG
Cyclo	47	SEQ. ID 0078	CUUCAUGAUCCAGGGCGGA
Cyclo	48	SEQ. ID 0079	UCAUGAUCCAGGGCGGAGA
Cyclo	49	SEQ. ID 0080	AUGAUCCAGGGCGGAGACU
Cyclo	50	SEQ. ID 0081	GAUCCAGGGCGGAGACUUC
Cyclo	51	SEQ. ID 0082	UCCAGGGCGGAGACUUCAC
Cyclo	52	SEQ. ID 0083	CAGGGCGGAGACUUCACCA
Cyclo	53	SEQ. ID 0084	GGGCGGAGACUUCACCAGG
Cyclo	54	SEQ. ID 0085	GCGGAGACUUCACCAGGGG
Cyclo	55	SEQ. ID 0086	GGAGACUUCACCAGGGGAG
Cyclo	56	SEQ. ID 0087	AGACUUCACCAGGGGAGAU
Cyclo	57	SEQ. ID 0088	ACUUCACCAGGGGAGAU
Cyclo	58	SEQ. ID 0089	UUCACCAGGGGAGAU
Cyclo	59	SEQ. ID 0090	CACCAGGGGAGAU
Cyclo	60	SEQ. ID 0091	CCAGGGGAGAU
Cyclo	61	SEQ. ID 0092	AGGGGAGAU
Cyclo	62	SEQ. ID 0093	GGGAGAU
Cyclo	63	SEQ. ID 0094	GAGAU
Cyclo	64	SEQ. ID 0095	GAUGGCACAGGAGGAA
Cyclo	65	SEQ. ID 0094	GAUGGCACAGGAGGAA
Cyclo	66	SEQ. ID 0096	UGGCACAGGAGGAA
Cyclo	67	SEQ. ID 0097	GCACAGGAGGAA
Cyclo	68	SEQ. ID 0098	ACAGGAGGAA
Cyclo	69	SEQ. ID 0099	AGGAGGAA
Cyclo	70	SEQ. ID 0100	GAGGAA
Cyclo	71	SEQ. ID 0101	GGAAAGAGCAUCUAC
Cyclo	72	SEQ. ID 0102	AAAGAGCAUCUAC
Cyclo	73	SEQ. ID 0103	AGAGCAUCUAC

Cyclo	74	SEQ. ID 0104	CAUCUACGGUGAGCGCUUC
Cyclo	75	SEQ. ID 0105	UCUACGGUGAGCGCUUCCC
Cyclo	76	SEQ. ID 0106	UACGGUGAGCGCUUCCCCG
Cyclo	77	SEQ. ID 0107	CGGUGAGCGCUUCCCCGAU
Cyclo	78	SEQ. ID 0108	GUGAGCGCUUCCCCGAUGA
Cyclo	79	SEQ. ID 0109	GAGCGCUUCCCCGAUGAGA
Cyclo	80	SEQ. ID 0110	GCGCUUCCCCGAUGAGAAC
Cyclo	81	SEQ. ID 0111	GCUUCCCCGAUGAGAACUU
Cyclo	82	SEQ. ID 0112	UUCCCCGAUGAGAACUJCA
Cyclo	83	SEQ. ID 0113	CCCCGAUGAGAACUUCAAA
Cyclo	84	SEQ. ID 0114	CCGAUGAGAACUUCAAACU
Cyclo	85	SEQ. ID 0115	GAUGAGAACUUCAAACUGA
Cyclo	86	SEQ. ID 0116	UGAGAACUUCAAACUGAAG
Cyclo	87	SEQ. ID 0117	AGAACUUCAAACUGAAGCA
Cyclo	88	SEQ. ID 0118	AACUUCAAACUGAAGCACU
Cyclo	89	SEQ. ID 0119	CUUCAAACUGAAGCACUAC
Cyclo	90	SEQ. ID 0120	UCAAACUGAAGCACUACGG
DB	1	SEQ. ID 0121	ACGGGCAAGGCCAAGUGGG
DB	2	SEQ. ID 0122	CGGGCAAGGCCAAGUGGGA
DB	3	SEQ. ID 0123	GGGCAAGGCCAAGUGGGAU
DB	4	SEQ. ID 0124	GGCAAGGCCAAGUGGGGAUG
DB	5	SEQ. ID 0125	GCAAGGCCAAGUGGGGAUGC
DB	6	SEQ. ID 0126	CAAGGCCAAGUGGGGAUGCC
DB	7	SEQ. ID 0127	AAGGCCAAGUGGGGAUGCCU
DB	8	SEQ. ID 0128	AGGCCAAGUGGGGAUGCCUG
DB	9	SEQ. ID 0129	GGCCAAGUGGGGAUGCCUGG
DB	10	SEQ. ID 0130	GCCAAGUGGGGAUGCCUGGA
DB	11	SEQ. ID 0131	CCAAGUGGGGAUGCCUGGAA
DB	12	SEQ. ID 0132	CAAGUGGGGAUGCCUGGAAU
DB	13	SEQ. ID 0133	AAGUGGGGAUGCCUGGAAUG
DB	14	SEQ. ID 0134	AGUGGGGAUGCCUGGAAUGA
DB	15	SEQ. ID 0135	GUGGGGAUGCCUGGAAUGAG
DB	16	SEQ. ID 0136	UGGGGAUGCCUGGAAUGAGC
DB	17	SEQ. ID 0137	GGGAUGCCUGGAAUGAGCU
DB	18	SEQ. ID 0138	GGAUGCCUGGAAUGAGCUG
DB	19	SEQ. ID 0139	GAUGCCUGGAAUGAGCUGA
DB	20	SEQ. ID 0140	AUGCCUGGAAUGAGCUGAA
DB	21	SEQ. ID 0141	UGCCUGGAAUGAGCUGAAA
DB	22	SEQ. ID 0142	GCCUGGAAUGAGCUGAAAG
DB	23	SEQ. ID 0143	CCUGGAAUGAGCUGAAAGG

DB	24	SEQ. ID 0144	CUGGAAUGAGCUGAAAGGG
DB	25	SEQ. ID 0145	UGGAAUGAGCUGAAAGGGA
DB	26	SEQ. ID 0146	GGAAUGAGCUGAAAGGGAC
DB	27	SEQ. ID 0147	GAAUGAGCUGAAAGGGACU
DB	28	SEQ. ID 0148	AAUGAGCUGAAAGGGACUU
DB	29	SEQ. ID 0149	AUGAGCUGAAAGGGACUUC
DB	30	SEQ. ID 0150	UGAGCUGAAAGGGACUUCC
DB	31	SEQ. ID 0151	GAGCUGAAAGGGACUUCCA
DB	32	SEQ. ID 0152	AGCUGAAAGGGACUUCCAA
DB	33	SEQ. ID 0153	GCUGAAAGGGACUUCCAAG
DB	34	SEQ. ID 0154	CUGAAAGGGACUUCCAAGG
DB	35	SEQ. ID 0155	UGAAAGGGACUUCCAAGGA
DB	36	SEQ. ID 0156	GAAAGGGACUUCCAAGGAA
DB	37	SEQ. ID 0157	AAAGGGACUUCCAAGGAAG
DB	38	SEQ. ID 0158	AAGGGACUUCCAAGGAAGA
DB	39	SEQ. ID 0159	AGGGACUUCCAAGGAAGAU
DB	40	SEQ. ID 0160	GGGACUUCCAAGGAAGAUG
DB	41	SEQ. ID 0161	GGACUUCCAAGGAAGAUGC
DB	42	SEQ. ID 0162	GACUUCCAAGGAAGAUGCC
DB	43	SEQ. ID 0163	ACUUCCAAGGAAGAUGCCA
DB	44	SEQ. ID 0164	CUUCCAAGGAAGAUGCCAU
DB	45	SEQ. ID 0165	UUCCAAGGAAGAUGCCAUG
DB	46	SEQ. ID 0166	UCCAAGGAAGAUGCCAUGA
DB	47	SEQ. ID 0167	CCAAGGAAGAUGCCAUGAA
DB	48	SEQ. ID 0168	CAAGGAAGAUGCCAUGAAA
DB	49	SEQ. ID 0169	AAGGAAGAUGCCAUGAAAG
DB	50	SEQ. ID 0170	AGGAAGAUGCCAUGAAAGC
DB	51	SEQ. ID 0171	GGAAGAUGCCAUGAAAGCU
DB	52	SEQ. ID 0172	GAAGAUGCCAUGAAAGCUU
DB	53	SEQ. ID 0173	AAGAUGCCAUGAAAGCUUA
DB	54	SEQ. ID 0174	AGAUGCCAUGAAAGCUUAC
DB	55	SEQ. ID 0175	GAUGCCAUGAAAGCUUACA
DB	56	SEQ. ID 0176	AUGCCAUGAAAGCUUACAU
DB	57	SEQ. ID 0177	UGCCAUGAAAGCUUACAUC
DB	58	SEQ. ID 0178	GCCAUGAAAGCUUACAUCA
DB	59	SEQ. ID 0179	CCAUGAAAGCUUACAUCAA
DB	60	SEQ. ID 0180	CAUGAAAGCUUACAUCAAC
DB	61	SEQ. ID 0181	AUGAAAGCUUACAUCAACA
DB	62	SEQ. ID 0182	UGAAAGCUUACAUCAACAA
DB	63	SEQ. ID 0183	GAAAGCUUACAUCAACAAA

DB	64	SEQ. ID 0184	AAAGCUUACAUCAACAAAG
DB	65	SEQ. ID 0185	AAGCUUACAUCAACAAAGU
DB	66	SEQ. ID 0186	AGCUUACAUCAACAAAGUA
DB	67	SEQ. ID 0187	GCUUACAUCAACAAAGUAG
DB	68	SEQ. ID 0188	CUUACAUCAACAAAGUAGA
DB	69	SEQ. ID 0189	UUACAUCAACAAAGUAGAA
DB	70	SEQ. ID 0190	UACAUCAACAAAGUAGAAG
DB	71	SEQ. ID 0191	ACAUCAACAAAGUAGAAGA
DB	72	SEQ. ID 0192	CAUCAACAAAGUAGAAGAG
DB	73	SEQ. ID 0193	AUCAACAAAGUAGAAGAGC
DB	74	SEQ. ID 0194	UCAACAAAGUAGAAGAGCU
DB	75	SEQ. ID 0195	CAACAAAGUAGAAGAGCUA
DB	76	SEQ. ID 0196	AACAAAGUAGAAGAGCUIA
DB	77	SEQ. ID 0197	ACAAAGUAGAAGAGCUIAAA
DB	78	SEQ. ID 0198	CAAAGUAGAAGAGCUIAAAG
DB	79	SEQ. ID 0199	AAAGUAGAAGAGCUIAAAGA
DB	80	SEQ. ID 0200	AAGUAGAAGAGCUIAAAGAA
DB	81	SEQ. ID 0201	AGUAGAAGAGCUIAAAGAAA
DB	82	SEQ. ID 0202	GUAGAAGAGCUIAAAGAAAA
DB	83	SEQ. ID 0203	UAGAAGAGCUIAAAGAAAAA
DB	84	SEQ. ID 0204	AGAAGAGCUIAAAGAAAAAA
DB	85	SEQ. ID 0205	GAAGAGCUIAAAGAAAAAAU
DB	86	SEQ. ID 0206	AAGAGCUIAAAGAAAAAAUA
DB	87	SEQ. ID 0207	AGAGCUIAAAGAAAAAAUAC
DB	88	SEQ. ID 0208	GAGCUIAAAGAAAAAAUACG
DB	89	SEQ. ID 0209	AGCUIAAAGAAAAAAUACGG
DB	90	SEQ. ID 0210	GCUIAAAGAAAAAAUACGGG
Luc	1	SEQ. ID 0211	AUCCUCAUAAAGGCCAAGA
Luc	2	SEQ. ID 0212	AGAUCCUCAUAAAGGCCAA
Luc	3	SEQ. ID 0213	AGAGAUCCUCAUAAAGGCC
Luc	4	SEQ. ID 0214	AGAGAGAUCCUCAUAAAGG
Luc	5	SEQ. ID 0215	UCAGAGAGAUCCUCAUAAA
Luc	6	SEQ. ID 0216	AAUCAGAGAGAUCCUCAUA
Luc	7	SEQ. ID 0217	AAAAUCAGAGAGAUCCUCA
Luc	8	SEQ. ID 0218	GAAAAUCAGAGAGAUCCU
Luc	9	SEQ. ID 0219	AAGAAAAUCAGAGAGAUCC
Luc	10	SEQ. ID 0220	GCAAGAAAAUCAGAGAGA
Luc	11	SEQ. ID 0221	ACGCAAGAAAAUCAGAGA
Luc	12	SEQ. ID 0222	CGACGCAAGAAAAUCAGA
Luc	13	SEQ. ID 0223	CUCGACGCAAGAAAAUCA

Luc	14	SEQ. ID 0224	AACUCGACGCAAGAAAAAU
Luc	15	SEQ. ID 0225	AAAACUCGACGCAAGAAAA
Luc	16	SEQ. ID 0226	GGAAAACUCGACGCAAGAA
Luc	17	SEQ. ID 0227	CCGAAAACUCGACGCAAG
Luc	18	SEQ. ID 0228	UACCGGAAAACUCGACGCA
Luc	19	SEQ. ID 0229	CUUACCGGAAAACUCGACG
Luc	20	SEQ. ID 0230	GUCUUACCGGAAAACUCGA
Luc	21	SEQ. ID 0231	AGGUCUUACCGGAAAACUC
Luc	22	SEQ. ID 0232	AAAGGUCUUACCGGAAAAC
Luc	23	SEQ. ID 0233	CGAAAGGUCUUACCGGAAA
Luc	24	SEQ. ID 0234	ACCGAAAGGUCUUACCGGA
Luc	25	SEQ. ID 0235	GUACCGAAAGGUCUUACCG
Luc	26	SEQ. ID 0236	AAGUACCGAAAGGUCUUAC
Luc	27	SEQ. ID 0237	CGAAGUACCGAAAGGUCUU
Luc	28	SEQ. ID 0238	GACGAAGUACCGAAAGGUC
Luc	29	SEQ. ID 0239	UGGACGAAGUACCGAAAGG
Luc	30	SEQ. ID 0240	UGUGGACGAAGUACCGAAA
Luc	31	SEQ. ID 0241	UUUGUGGACGAAGUACCGA
Luc	32	SEQ. ID 0242	UGUUUGUGGACGAAGUACC
Luc	33	SEQ. ID 0243	UGUGUUUGUGGACGAAGUA
Luc	34	SEQ. ID 0244	GUUGUGUUUGUGGACGAAG
Luc	35	SEQ. ID 0245	GAGUUGUGUUUGUGGACGA
Luc	36	SEQ. ID 0246	AGGAGUUGUGUUUGUGGAC
Luc	37	SEQ. ID 0247	GGAGGAGUUGUGUUUGUGG
Luc	38	SEQ. ID 0248	GCGGAGGAGUUGUGUUUGU
Luc	39	SEQ. ID 0249	GCGCGGAGGAGUUGUGUUU
Luc	40	SEQ. ID 0250	UUGCGCGGAGGAGUUGUGU
Luc	41	SEQ. ID 0251	AGUUGCGCGGAGGAGUUGU
Luc	42	SEQ. ID 0252	AAAGUUGCGCGGAGGAGUU
Luc	43	SEQ. ID 0253	AAAAAGUUGCGCGGAGGAG
Luc	44	SEQ. ID 0254	CGAAAAAGUUGCGCGGAGG
Luc	45	SEQ. ID 0255	CGCGAAAAAGUUGCGCGGA
Luc	46	SEQ. ID 0256	ACCGCGAAAAAGUUGCGCG
Luc	47	SEQ. ID 0257	CAACCGCGAAAAAGUUGCG
Luc	48	SEQ. ID 0258	AACAACCGCGAAAAAGUUG
Luc	49	SEQ. ID 0259	GUAACAACCGCGAAAAAGU
Luc	50	SEQ. ID 0260	AAGUAACAACCGCGAAAAA
Luc	51	SEQ. ID 0261	UCAAGUAACAACCGCGAAA
Luc	52	SEQ. ID 0262	AGUCAAGUAACAACCGCGA
Luc	53	SEQ. ID 0263	CCAGUCAAGUAACAACCGC

Luc	54	SEQ. ID 0264	CGCCAGUCAAGUAACAACC
Luc	55	SEQ. ID 0265	GUCGCCAGUCAAGUAACAA
Luc	56	SEQ. ID 0266	ACGUCGCCAGUCAAGUAAC
Luc	57	SEQ. ID 0267	UUACGUCGCCAGUCAAGUA
Luc	58	SEQ. ID 0268	GAUUACGUCGCCAGUCAAG
Luc	59	SEQ. ID 0269	UGGAUUACGUCGCCAGUCA
Luc	60	SEQ. ID 0270	CGUGGAUUACGUCGCCAGU
Luc	61	SEQ. ID 0271	AUCGUGGAUUACGUCGCCA
Luc	62	SEQ. ID 0272	AGAUCGUGGAUUACGUCGC
Luc	63	SEQ. ID 0273	AGAGAUCGUGGAUUACGUC
Luc	64	SEQ. ID 0274	AAAGAGAUCGUGGAUUACG
Luc	65	SEQ. ID 0275	AAAAAGAGAUCGUGGAUUA
Luc	66	SEQ. ID 0276	GGAAAAAGAGAUCGUGGAU
Luc	67	SEQ. ID 0277	ACGGAAAAAGAGAUCGUGG
Luc	68	SEQ. ID 0278	UGACGGAAAAAGAGAUCGU
Luc	69	SEQ. ID 0279	GAUGACGGAAAAAGAGAUC
Luc	70	SEQ. ID 0280	ACGAUGACGGAAAAAGAGA
Luc	71	SEQ. ID 0281	AGACGAUGACGGAAAAAGA
Luc	72	SEQ. ID 0282	AAAGACGAUGACGGAAAAA
Luc	73	SEQ. ID 0283	GGAAAGACGAUGACGGAAA
Luc	74	SEQ. ID 0284	ACGGAAAGACGAUGACGGA
Luc	75	SEQ. ID 0285	GCACGGAAAGACGAUGACG
Luc	76	SEQ. ID 0286	GAGCACGGAAAGACGAUGA
Luc	77	SEQ. ID 0287	UGGAGCACGGAAAGACGAU
Luc	78	SEQ. ID 0288	UUUGGAGCACGGAAAGACG
Luc	79	SEQ. ID 0289	GUUUUGGAGCACGGAAAGA
Luc	80	SEQ. ID 0290	UUGUUUUGGAGCACGGAAA
Luc	81	SEQ. ID 0291	UGUUGUUUUGGAGCACGGA
Luc	82	SEQ. ID 0292	GUUGUUGUUUUGGAGCACG
Luc	83	SEQ. ID 0293	CCGUUGUUGUUUUGGAGCA
Luc	84	SEQ. ID 0294	CGCCGUUGUUGUUUUGGAG
Luc	85	SEQ. ID 0295	GCCGCCGUUGUUGUUUUGG
Luc	86	SEQ. ID 0296	CCGCCGCCGUUGUUGUUUU
Luc	87	SEQ. ID 0297	UCCCGCCGCCGUUGUUGUU
Luc	88	SEQ. ID 0298	CUUCCCGCCGCCGUUGUUG
Luc	89	SEQ. ID 0299	AACUCCCGCCGCCGUUGU
Luc	90	SEQ. ID 0300	UGAACUCCCGCCGCCGUU

Example II. Validation of the Algorithm using DBI, Luciferase, PLK, EGFR, and SEAP

The algorithm (Formula VIII) identified siRNAs for five genes, human DBI, firefly luciferase (fLuc), renilla luciferase (rLuc), human PLK, and human secreted alkaline phosphatase (SEAP). Four individual siRNAs were selected on the basis of their SMARTscores™ derived by analysis of their sequence using Formula VIII (all of the siRNAs would be selected with Formula IX as well) and analyzed for their ability to silence their targets' expression. In addition to the scoring, a BLAST search was conducted for each siRNA. To minimize the potential for off-target silencing effects, only those target sequences with more than three mismatches against unrelated sequences were selected. Semizarov, *et al*, *Specificity of short interfering RNA determined through gene expression signatures*. Proc. Natl. Acad. Sci. U.S.A. 2003, 100:6347. These duplexes were analyzed individually and in pools of 4 and compared with several siRNAs that were randomly selected. The functionality was measured a percentage of targeted gene knockdown as compared to controls. All siRNAs were transfected as described by the methods above at 100 nM concentration into HEK293 using Lipofectamine 2000. The level of the targeted gene expression was evaluated by B-DNA as described above and normalized to the non-specific control. **Figure 10** shows that the siRNAs selected by the algorithm disclosed herein were significantly more potent than randomly selected siRNAs. The algorithm increased the chances of identifying an F50 siRNA from 48% to 91%, and an F80 siRNA from 13% to 57%. In addition, pools of SMART siRNA silence the selected target better than randomly selected pools (see Figure 10F).

Example III. Validation of the Algorithm Using Genes Involved in Clathrin-Dependent Endocytosis.

Components of clathrin-mediated endocytosis pathway are key to modulating intracellular signaling and play important roles in disease. Chromosomal rearrangements that result in fusion transcripts between the Mixed-Lineage Leukemia gene (MLL) and CALM (Clathrin assembly lymphoid myeloid leukemia gene) are believed to play a role in leukemogenesis. Similarly, disruptions in Rab7 and Rab9, as well as HIP1 (Huntingtin-interacting protein), genes that are believed to be involved in endocytosis, are potentially responsible for ailments resulting in lipid storage, and neuronal diseases, respectively. For these reasons, siRNA directed

against clathrin and other genes involved in the clathrin-mediated endocytotic pathway are potentially important research and therapeutic tools.

siRNAs directed against genes involved in the clathrin-mediated endocytosis pathways were selected using Formula VIII. The targeted genes were clathrin heavy chain (CHC, accession # NM_004859), clathrin light chain A (CLCa, NM_001833), clathrin light chain B (CLCb, NM_001834), CALM (U45976), β 2 subunit of AP-2 (β 2, NM_001282), Eps15 (NM_001981), Eps15R (NM_021235), dynamin II (DYNII, NM_004945), Rab5a (BC001267), Rab5b (NM_002868), Rab5c (AF141304), and EEA.1 (XM_018197).

For each gene, four siRNAs duplexes with the highest scores were selected and a BLAST search was conducted for each of them using the Human EST database. In order to minimize the potential for off-target silencing effects, only those sequences with more than three mismatches against un-related sequences were used. All duplexes were synthesized at Dharmacon, Inc. as 21-mers with 3'-UU overhangs using a modified method of 2'-ACE chemistry Scaringe, *Advanced 5'-silyl-2'-orthoester approach to RNA oligonucleotide synthesis*, Methods Enzymol 2000, 317:3 and the antisense strand was chemically phosphorylated to insure maximized activity.

HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, antibiotics and glutamine. siRNA duplexes were resuspended in 1X siRNA Universal buffer (Dharmacon, Inc.) to 20 μ M prior to transfection. HeLa cells in 12-well plates were transfected twice with 4 μ l of 20 μ M siRNA duplex in 3 μ l Lipofectamine 2000 reagent (Invitrogen, Carlsbad, California, USA) at 24-hour intervals. For the transfections in which 2 or 3 siRNA duplexes were included, the amount of each duplex was decreased, so that the total amount was the same as in transfections with single siRNAs. Cells were plated into normal culture medium 12 hours prior to experiments, and protein levels were measured 2 or 4 days after the first transfection.

Equal amounts of lysates were resolved by electrophoresis, blotted, and stained with the antibody specific to targeted protein, as well as antibodies specific to

- unrelated proteins, PP1 phosphatase and Tsg101 (not shown). The cells were lysed in Triton X-100/glycerol solubilization buffer as described previously. Tebar, Bohlander, & Sorkin, *Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-mediated Traffic*, Mol. Biol. Cell Aug 1999, 10:2687. Cell lysates were electrophoresed, transferred to nitrocellulose membranes, and Western blotting was performed with several antibodies followed by detection using enhanced chemiluminescence system (Pierce, Inc). Several x-ray films were analyzed to determine the linear range of the chemiluminescence signals, and the quantifications were performed using densitometry and Alphamager v5.5 software (Alpha Innotech Corporation). In experiments with Eps15R-targeted siRNAs, cell lysates were subjected to immunoprecipitation with Ab860, and Eps15R was detected in immunoprecipitates by Western blotting as described above.
- 15 The antibodies to assess the levels of each protein by Western blot were obtained from the following sources: monoclonal antibody to clathrin heavy chain (TD.1) was obtained from American Type Culture Collection (Rockville, MD, USA); polyclonal antibody to dynamin II was obtained from Affinity Bioreagents, Inc. (Golden, CO, USA); monoclonal antibodies to EEA.1 and Rab5a were purchased from BD Transduction Laboratories (Los Angeles, CA, USA); the monoclonal antibody to Tsg101 was purchased from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA, USA); the monoclonal antibody to GFP was from ZYMED Laboratories Inc. (South San Francisco, CA, USA); the rabbit polyclonal antibodies Ab32 specific to α -adaptins and Ab20 to CALM were described previously Sorkin, *et al*, *Stoichiometric Interaction of the Epidermal Growth Factor Receptor with the Clathrin-associated Protein Complex AP-2*, J. Biol. Chem. Jan 1995, 270:619, the polyclonal antibodies to clathrin light chains A and B were kindly provided by Dr. F. Brodsky (UCSF); monoclonal antibodies to PP1 (BD Transduction Laboratories) and α -Actinin (Chemicon) were kindly provided by Dr. M. Dell'Acqua (University of Colorado); Eps15 Ab577 and Eps15R Ab860 were kindly provided by Dr. P.P. Di Fiore (European Cancer Institute).

Figure 11 demonstrates the *in vivo* functionality of 48 individual siRNAs, selected using Formula VIII (most of them will meet the criteria incorporated by Formula IX as well) targeting 12 genes. Various cell lines were transfected with siRNA duplexes (*Dup1-4*) or pools of siRNA duplexes (Pool), and the cells were

5 lysed 3 days after transfection with the exception of CALM (2 days) and $\beta 2$ (4 days).

Note a $\beta 1$ -adaptin band (part of AP-1 Golgi adaptor complex) that runs slightly slower than $\beta 2$ adaptin. CALM has two splice variants, 66 and 72 kD. The full-length Eps15R (a doublet of ~ 130 kD) and several truncated spliced forms of \sim

10 100 kD and ~ 70 kD were detected in Eps15R immunoprecipitates (shown by arrows). The cells were lysed 3 days after transfection. Equal amounts of lysates were resolved by electrophoresis and blotted with the antibody specific to a targeted protein (GFP antibody for YFP fusion proteins) and the antibody specific to unrelated

15 proteins PP1 phosphatase or α -actinin, and TSG101. The amount of protein in each specific band was normalized to the amount of non-specific proteins in each lane of the gel. Nearly all of them appear to be functional, which establishes that Formula VIII and IX can be used to predict siRNAs' functionality in general in a genome wide manner.

20 To generate the fusion of yellow fluorescent protein (YFP) with Rab5b or Rab5c (YFP-Rab5b or YFP-Rab5c), a DNA fragment encoding the full-length human Rab5b or Rab5c was obtained by PCR using *Pfu* polymerase (Stratagene) with a *SacI* restriction site introduced into the 5' end and a *KpnI* site into the 3' end and cloned into pEYFP-C1 vector (CLONTECH, Palo Alto, CA, USA). GFP-CALM and YFP-

25 Rab5a were described previously Tebar, Bohlander, & Sorkin, *Clathrin Assembly Lymphoid Myeloid Leukemia (CALM) Protein: Localization in Endocytic-coated Pits, Interactions with Clathrin, and the Impact of Overexpression on Clathrin-mediated Traffic*, Mol. Biol. Cell Aug 1999, 10:2687.

30 **Example III. Validation of the Algorithm Using Eg5, GADPH, ATE1, MEK2, MEK1, QB, LaminA/C, c-myc, human cyclophilin, and mouse cyclophilin.**

A number of genes have been identified as playing potentially important roles in disease etiology. Expression profiles of normal and diseased kidneys has implicated Edg5 in immunoglobulin A neuropathy, a common renal glomerular disease. Myc1, MEK1/2 and other related kinases have been associated with one or more cancers, while lamins have been implicated in muscular dystrophy and other diseases. For these reasons, siRNA directed against the genes encoding these classes of molecules would be important research and therapeutic tools.

Figure 12 illustrates four siRNAs targeting 10 different genes (**Table V** for sequence and accession number information) that were selected according to the Formula VIII and assayed as individuals and pools in HEK293 cells. The level of siRNA induced silencing was measured using the B-DNA assay. These studies demonstrated that thirty-six out of the forty individual SMART-selected siRNA tested are functional (90%) and all 10 pools are fully functional.

Example V. Validation of the Algorithm Using Bcl2

Bcl-2 is a ~25kD, 205-239 amino acid, anti-apoptotic protein that contains considerable homology with other members of the BCL family including BCLX, MCL1, BAX, BAD, and BIK. The protein exists in at least two forms (Bcl2a, which has a hydrophobic tail for membrane anchorage, and Bcl2b, which lacks the hydrophobic tail) and is predominantly localized to the mitochondrial membrane. While Bcl2 expression is widely distributed, particular interest has focused on the expression of this molecule in B₁ and T cells. Bcl2 expression is down-regulated in normal germinal center B cells yet in a high percentage of follicular lymphomas, Bcl2 expression has been observed to be elevated. Cytological studies have identified a common translocation ((14;18)(q32;q32)) amongst a high percentage (>70%) of these lymphomas. This genetic lesion places the Bcl2 gene in juxtaposition to immunoglobulin heavy chain gene (IgH) encoding sequences and is believed to enforce inappropriate levels of gene expression, and resistance to programmed cell death in the follicle center B cells. In other cases, hypomethylation of the Bcl2 promoter leads to enhanced expression and again, inhibition of apoptosis. In addition to cancer, dysregulated expression of Bcl-2 has been correlated with multiple sclerosis and various neurological diseases.

The correlation between Bcl-2 translocation and cancer makes this gene an attractive target for RNAi. Identification of siRNA directed against the bcl2 transcript (or Bcl2-IgH fusions) would further our understanding Bcl2 gene function and possibly provide a future therapeutic agent to battle diseases that result from altered expression or function of this gene.

In Silico Identification of Functional siRNA

To identify functional and hyperfunctional siRNA against the Bcl2 gene, the sequence for Bcl-2 was downloaded from the NCBI Unigene database and analyzed using the Formula VIII algorithm. As a result of these procedures, both the sequence and SMARTscores™ of the Bcl2 siRNA were obtained and ranked according to their functionality. Subsequently, these sequences were BLAST'ed (database) to insure that the selected sequences were specific and contained minimal overlap with unrelated genes. The SMARTscores™ for the top 10 Bcl-2 siRNA are identified in Figure 13.

In Vivo Testing of Bcl-2 SiRNA

Bcl-2 siRNAs having the top ten SMARTscores™ were selected and tested in a functional assay to determine silencing efficiency. To accomplish this, each of the ten duplexes were synthesized using 2'-O-ACE chemistry and transfected at 100nM concentrations into cells. Twenty-four hours later assays were performed on cell extracts to assess the degree of target silencing. Controls used in these experiments included mock transfected cells, and cells that were transfected with a non-specific siRNA duplex.

The results of these experiments are presented below (and in Figure 14) and show that all ten of the selected siRNA induce 80% or better silencing of the Bcl2 message at 100nM concentrations. These data verify that the algorithm successfully identified functional Bcl2 siRNA and provide a set of functional agents that can be used in experimental and therapeutic environments.

5	siRNA 1	GGGAGAUAGUGAUGAAGUA	SEQ. ID NO. 301
	siRNA 2	GAAGUACAUCCAUUAUAAG	SEQ. ID NO. 302
	siRNA 3	GUACGACAACCGGGAGAU	SEQ. ID NO. 303
	siRNA 4	AGAUAGUGAUGAAGUACAU	SEQ. ID NO. 304
	siRNA 5	UGAAGACUCUGCUCAGUUU	SEQ. ID NO. 305
	siRNA 6	GCAUGCGGCCUCUGUUUGA	SEQ. ID NO. 306
	siRNA 7	UGCGGCCUCUGUUUGAUUU	SEQ. ID NO. 307
	siRNA 8	GAGAUAGUGAUGAAGUACA	SEQ. ID NO. 308
	siRNA 9	GGAGAUAGUGAUGAAGUAC	SEQ. ID NO. 309
	siRNA 10	GAAGACUCUGCUCAGUUUG	SEQ. ID NO. 310

10 Bcl2 siRNA: Sense Strand, 5'→3'

Example VI. Sequences Selected by the Algorithm

Sequences of the siRNAs selected using Formulas (Algorithms) VIII and IX
 15 with their corresponding ranking, which have been evaluated for the silencing activity *in vivo* in the present study (Formula VIII and IX, respectively).

TABLE V

Gene Name	Accession Number	SEQ. ID NO.	FTIISeqTence	Formula	Formula
				VIII	IX
CLTC	NM_004859	SEQ. ID NO. 0301	GAAAGAATCTGTAGAGAAA	76	94.2
CLTC	NM_004859	SEQ. ID NO. 0302	GCAATGAGCTGTTTGAAGA	65	39.9
CLTC	NM_004859	SEQ. ID NO. 0303	TGACAAAGGTGGATAAATT	57	38.2
CLTC	NM_004859	SEQ. ID NO. 0304	GGAAATGGATCTCTTTGAA	54	49.4
CLTA	NM_001833	SEQ. ID NO. 0305	GGAAAGTAATGGTCCAACA	22	55.5
CLTA	NM_001833	SEQ. ID NO. 0306	AGACAGTTATGCAGCTATT	4	22.9
CLTA	NM_001833	SEQ. ID NO. 0307	CCAATTCTCGGAAGCAAGA	1	17
CLTA	NM_001833	SEQ. ID NO. 0308	GAAAGTAATGGTCCAACAG	-1	-13
CLTB	NM_001834	SEQ. ID NO. 0309	GCGCCAGAGTGAACAAGTA	17	57.5
CLTB	NM_001834	SEQ. ID NO. 0310	GAAGGTGGCCCAGCTATGT	15	-8.6
CLTB	NM_001834	SEQ. ID NO. 0311	GGAACCAGCGCCAGAGTGA	13	40.5
CLTB	NM_001834	SEQ. ID NO. 0312	GAGCGAGATTGCAGGCATA	20	61.7
CALM	U45976	SEQ. ID NO. 0313	GTTAGTATCTGATGACTTG	36	-34.6
CALM	U45976	SEQ. ID NO. 0314	GAAATGGAACCACTAAGAA	33	46.1
CALM	U45976	SEQ. ID NO. 0315	GGAAATGGAACCACTAAGA	30	61.2
CALM	U45976	SEQ. ID NO. 0316	CAACTACACTTTCCAATGC	28	6.8
EPS15	NM_001981	SEQ. ID NO. 0317	CCACCAAGATTTTCATGATA	48	25.2

EPS15	NM_001981	SEQ. ID NO. 0318	GATCGGAACCTCCAACAAGA	43	49.3
EPS15	NM_001981	SEQ. ID NO. 0319	AAACGGAGCTACAGATTAT	39	11.5
EPS15	NM_001981	SEQ. ID NO. 0320	CCACACAGCATTCTTGTA	33	-23.6
EPS15R	NM_021235	SEQ. ID NO. 0321	GAAGTTACCTTGAGCAATC	48	33
EPS15R	NM_021235	SEQ. ID NO. 0322	GGACTTGCCGATCCAGAA	27	33
EPS15R	NM_021235	SEQ. ID NO. 0323	GCACTTGATCGAGATGAG	20	1.3
EPS15R	NM_021235	SEQ. ID NO. 0324	CAAAGACCAATTCGCGTTA	17	27.7
DNM2	NM_004945	SEQ. ID NO. 0325	CCGAATCAATCGCATCTTC	6	-29.6
DNM2	NM_004945	SEQ. ID NO. 0326	GACATGATCCTGCAGTTCA	5	-14
DNM2	NM_004945	SEQ. ID NO. 0327	GAGCGAATCGTCACCACTT	5	24
DNM2	NM_004945	SEQ. ID NO. 0328	CCTCCGAGCTGGCGTCTAC	-4	-63.6
ARF6	AF93885	SEQ. ID NO. 0329	TCACATGGTTAACCTCTAA	27	-21.1
ARF6	AF93885	SEQ. ID NO. 0330	GATGAGGGACGCCATAATC	7	-38.4
ARF6	AF93885	SEQ. ID NO. 0331	CCTCTAACTACAAATCTTA	4	16.9
ARF6	AF93885	SEQ. ID NO. 0332	GGAAGGTGCTATCCAAAAT	4	11.5
RAB5A	BC001267	SEQ. ID NO. 0333	GCAAGCAAGTCCTAACATT	40	25.1
RAB5A	BC001267	SEQ. ID NO. 0334	GGAAGAGGAGTAGACCTTA	17	50.1
RAB5A	BC001267	SEQ. ID NO. 0335	AGGAATCAGTGTGTAGTA	16	11.5
RAB5A	BC001267	SEQ. ID NO. 0336	GAAGAGGAGTAGACCTTAC	12	7
RAB5B	NM_002868	SEQ. ID NO. 0337	GAAAGTCAAGCCTGGTATT	14	18.1
RAB5B	NM_002868	SEQ. ID NO. 0338	AAAGTCAAGCCTGGTATTA	6	-17.8
RAB5B	NM_002868	SEQ. ID NO. 0339	GCTATGAACGTGAATGATC	3	-21.1
RAB5B	NM_002868	SEQ. ID NO. 0340	CAAGCCTGGTATTACGTTT	-7	-37.5
RAB5C	AF141304	SEQ. ID NO. 0341	GGAACAAGATCTGTCAATT	38	51.9
RAB5C	AF141304	SEQ. ID NO. 0342	GCAATGAACGTGAACGAAA	29	43.7
RAB5C	AF141304	SEQ. ID NO. 0343	CAATGAACGTGAACGAAAT	18	43.3
RAB5C	AF141304	SEQ. ID NO. 0344	GGACAGGAGCGGTATCACA	6	18.2
EEA1	XM_018197	SEQ. ID NO. 0345	AGACAGAGCTTGAGAATAA	67	64.1
EEA1	XM_018197	SEQ. ID NO. 0346	GAGAAGATCTTTATGCAA	60	48.7
EEA1	XM_018197	SEQ. ID NO. 0347	GAAGAGAAATCAGCAGATA	58	45.7
EEA1	XM_018197	SEQ. ID NO. 0348	GCAAGTAACTCAACTAACA	56	72.3
AP2B1	NM_001282	SEQ. ID NO. 0349	GAGCTAATCTGCCACATTG	49	-12.4
AP2B1	NM_001282	SEQ. ID NO. 0350	GCAGATGAGTTACTAGAAA	44	48.9
AP2B1	NM_001282	SEQ. ID NO. 0351	CAACTTAATTGTCCAGAAA	41	28.2
AP2B1	NM_001282	SEQ. ID NO. 0352	CAACACAGGATTCTGATAA	33	-5.8
PLK	NM_005030	SEQ. ID NO. 0353	AGATTGTGCCTAAGTCTCT	-35	-3.4
PLK	NM_005030	SEQ. ID NO. 0354	ATGAAGATCTGGAGGTGAA	0	-4.3
PLK	NM_005030	SEQ. ID NO. 0355	TTTGAGACTTCTTGCCTAA	-5	-27.7
PLK	NM_005030	SEQ. ID NO. 0356	AGATCACCTCCTTAAATA	15	72.3
GAPDH	NM_002046	SEQ. ID NO. 0357	CAACGGATTGTCGTATT	27	-2.8

GAPDH	NM_002046	SEQ. ID NO. 0358	GAAATCCCATCACCATCTT	24	3.9
GAPDH	NM_002046	SEQ. ID NO. 0359	GACCTCAACTACATGGTTT	22	-22.9
GAPDH	NM_002046	SEQ. ID NO. 0360	TGGTTTACATGTTCCAATA	9	9.8
c-Myc		SEQ. ID NO. 0361	GAAGAAATCGATGTTGTTT	31	-11.7
c-Myc		SEQ. ID NO. 0362	ACACAAACTTGAACAGCTA	22	51.3
c-Myc		SEQ. ID NO. 0363	GGAAGAAATCGATGTTGTT	18	26
c-Myc		SEQ. ID NO. 0364	GAAACGACGAGAACAGTTG	18	-8.9
MAP2K1	NM_002755	SEQ. ID NO. 0365	GCACATGGATGGAGGTTCT	26	16
MAP2K1	NM_002755	SEQ. ID NO. 0366	GCAGAGAGAGCAGATTTGA	16	0.4
MAP2K1	NM_002755	SEQ. ID NO. 0367	GAGGTTCTCTGGATCAAGT	14	15.5
MAP2K1	NM_002755	SEQ. ID NO. 0368	GAGCAGATTTGAAGCAACT	14	18.5
MAP2K2	NM_030662	SEQ. ID NO. 0369	CAAAGACGATGACTTCGAA	37	26.4
MAP2K2	NM_030662	SEQ. ID NO. 0370	GATCAGCATTTGCATGGAA	24	-0.7
MAP2K2	NM_030662	SEQ. ID NO. 0371	TCCAGGAGTTTGTCAATAA	17	-4.5
MAP2K2	NM_030662	SEQ. ID NO. 0372	GGAAGCTGATCCACCTTGA	16	59.2
KNLS1(EG5)	NM_004523	SEQ. ID NO. 0373	GCAGAAATCTAAGGATATA	53	35.8
KNLS1(EG5)	NM_004523	SEQ. ID NO. 0374	CAACAAGGATGAAGTCTAT	50	18.3
KNLS1(EG5)	NM_004523	SEQ. ID NO. 0375	CAGCAGAAATCTAAGGATA	41	32.7
KNLS1(EG5)	NM_004523	SEQ. ID NO. 0376	CTAGATGGCTTTCTCAGTA	39	3.9
CyclophilinA	NM_021130	SEQ. ID NO. 0377	AGACAAGGTCCCAAAGACA	-16	58.1
CyclophilinA	NM_021130	SEQ. ID NO. 0378	GGAATGGCAAGACCAGCAA	-6	36
CyclophilinA	NM_021130	SEQ. ID NO. 0379	AGAATTATTCCAGGGTTTA	-3	16.1
CyclophilinA	NM_021130	SEQ. ID NO. 0380	GCAGACAAGGTCCCAAAGA	8	8.9
LAMIN A/C	NM_170707	SEQ. ID NO. 0381	AGAAGCAGCTTCAGGATGA	31	38.8
LAMIN A/C	NM_170707	SEQ. ID NO. 0382	GAGCTTGACTTCCAGAAGA	33	22.4
LAMIN A/C	NM_170707	SEQ. ID NO. 0383	CCACCGAAGTTCACCCTAA	21	27.5
LAMIN A/C	NM_170707	SEQ. ID NO. 0384	GAGAAGAGCTCCTCCATCA	55	30.1
CyclophilinB	M60857	SEQ. ID NO. 0385	GAAAGAGCATCTACGGTGA	41	83.9
CyclophilinB	M60857	SEQ. ID NO. 0386	GAAAGGATTTGGCTACAAA	53	59.1
CyclophilinB	M60857	SEQ. ID NO. 0387	ACAGCAAATTCATCGTGT	-20	28.8
CyclophilinB	M60857	SEQ. ID NO. 0388	GGAAAGACTGTTCCAAAAA	2	27
DBI1	NM_020548	SEQ. ID NO. 0389	CAACACGCCTCATCCTCTA	27	-7.6
DBI2	NM_020548	SEQ. ID NO. 0390	CATGAAAGCTTACATCAAC	25	-30.8
DBI3	NM_020548	SEQ. ID NO. 0391	AAGATGCCATGAAAGCTTA	17	22
DBI4	NM_020548	SEQ. ID NO. 0392	GCACATACCGCTGAGTCT	15	3.9
rLUC1		SEQ. ID NO. 0393	GATCAAATCTGAAGAAGGA	57	49.2
rLUC2		SEQ. ID NO. 0394	GCCAAGAAGTTTCCTAATA	50	13.7
rLUC3		SEQ. ID NO. 0395	CAGCATATCTTGAACCATT	41	-2.2
rLUC4		SEQ. ID NO. 0396	GAACAAAGGAAACGGATGA	39	29.2
SeAP1	NM_031313	SEQ. ID NO. 0397	CGGAAACGGTCCAGGCTAT	6	26.9

SeAP2	NM_031313	SEQ. ID NO. 0398	GCTTCGAGCAGACATGATA	4	-11.2
SeAP3	NM_031313	SEQ. ID NO. 0399	CCTACACGGTCCTCCTATA	4	4.9
SeAP4	NM_031313	SEQ. ID NO. 0400	GCCAAGAACCTCATCATCT	1	-9.9
fLUC1		SEQ. ID NO. 0401	GATATGGGCTGAATACAAA	54	40.4
fLUC2		SEQ. ID NO. 0402	GCACTCTGATTGACAAATA	47	54.7
fLUC3		SEQ. ID NO. 0403	TGAAGTCTCTGATTAAGTA	46	34.5
fLUC4		SEQ. ID NO. 0404	TCAGAGAGATCCTCATAAA	40	11.4
mCyclo_1	NM_008907	SEQ. ID NO. 0405	GCAAGAAGATCACCATTTC	52	46.4
mCyclo_2	NM_008907	SEQ. ID NO. 0406	GAGAGAAATTTGAGGATGA	36	70.7
mCyclo_3	NM_008907	SEQ. ID NO. 0407	GAAAGGATTTGGCTATAAG	35	-1.5
mCyclo_4	NM_008907	SEQ. ID NO. 0408	GAAAGAAGGCATGAACATT	27	10.3
BCL2_1	NM_000633	SEQ. ID NO. 0409	GGGAGATAGTGATGAAGTA	21	72
BCL2_2	NM_000633	SEQ. ID NO. 0410	GAAGTACATCCATTATAAG	1	3.3
BCL2_3	NM_000633	SEQ. ID NO. 0411	GTACGACAACCGGGAGATA	1	35.9
BCL2_4	NM_000633	SEQ. ID NO. 0412	AGATAGTGATGAAGTACAT	-12	22.1
BCL2_5	NM_000633	SEQ. ID NO. 0413	TGAAGACTCTGCTCAGTTT	36	19.1
BCL2_6	NM_000633	SEQ. ID NO. 0414	GCATGCGGCCTCTGTTTGA	5	-9.7
QB1	NM_003365.1	SEQ. ID NO. 0415	GCACACAGCUUACUACAUC	52	-4.8
QB2	NM_003365.1	SEQ. ID NO. 0416	GAAAUGCCUGGUAUCUCA	49	22.1
QB3	NM_003365.1	SEQ. ID NO. 0417	GAAGGAACGUGAUGUGAUC	34	22.9
QB4	NM_003365.1	SEQ. ID NO. 0418	GCACUACUCCUGUGUGUGA	28	20.4
ATE1-1	NM_007041	SEQ. ID NO. 0419	GAACCCAGCUGGAGAACUU	45	15.5
ATE1-2	NM_007041	SEQ. ID NO. 0420	GAUUAUACAGUGUGAUCUUA	40	12.2
ATE1-3	NM_007041	SEQ. ID NO. 0421	GUACUACGAUCCUGAUUAU	37	32.9
ATE1-4	NM_007041	SEQ. ID NO. 0422	GUGCCGACCUUUACAAUUU	35	18.2
EGFR-1	NM_005228	SEQ. ID NO. 0423	GAAGGAAACTGAATTCAAA	68	79.4
EGFR-1	NM_005228	SEQ. ID NO. 0424	GGAAATATGTACTACGAAA	49	49.5
EGFR-1	NM_005228	SEQ. ID NO. 0425	CCACAAAGCAGTGAATTTA	41	7.6
EGFR-1	NM_005228	SEQ. ID NO. 0426	GTAACAAGCTCACGCAGTT	40	25.9

Example VII. Genome-Wide Application of the Algorithm

The examples described above demonstrate that the algorithm(s) can

5 successfully identify functional siRNA and that these duplexes can be used to induce the desirable phenotype of transcriptional knockdown or knockout. Each gene or family of genes in each organism plays an important role in maintaining physiological homeostasis and the algorithm can be used to develop functional, highly functional, or hyperfunctional siRNA to each gene. To accomplish this for the human genome, the

10 entire online ncbi refseq database was accessed through Entrez (efetch). The database

was processed through Formula VIII. For each gene the top 80 –100 scores for siRNAs were obtained and BLAST'ed to insure that the selected sequences are specific in targeting the gene of choice. These sequences are provided on the enclosed CDs in electronic form. Accordingly, Applicants hereby incorporate by reference the material submitted herewith, in duplicate on the compact disks labeled COPY 1 – TABLES PART, DISK 1/1, TABLES 12 –15, Filed with RO/US under PCT AI sec. 801(a), Operating System: MS-Windows, COPY 2 – TABLES PART, DISK 1/1, TABLES 12 –15, Filed with RO/US under PCT AI sec. 801(a), Operating System: MS-Windows, COPY 3 – TABLES PART, DISK 1/1, TABLES 12 –15, Filed with RO/US under PCT AI sec. 801(a), Operating System: MS-Windows,; which copies are identical, in files entitled Table_12.txt, date of creation June 26, 2003, with a size of 31,045 kb; Table_13.txt; date of creation November 13, 2003, with a size of 78,451 kb; Table_14.txt, date of creation November 13, 2003, with a size of 454 kb; and Table_15.txt date of creation November 13, 2003, with a size of 1,690 kb.

With respect to the disks, there are four tables on each disk copy in text format: Tables XII –XV. Table XII, which is located in a file entitled Table_12.txt, provides a list of the 80-100 sequences for each target, identified by Formula VIII as having the highest relative SMARTscoresTM for the target analyzed. Table XIII, which is located in a file entitled Table_13.txt, provides the SMARTscoresTM, and for each gene, a pool pick of up to four sequences is denoted. (The denotation of "1" in Table XIII means that it is a pool pick.) These pool pick sequences represent the most functional siRNAs for the corresponding target. Any 1, 2, 3, or 4 of the pool pick sequences could be used for gene silencing. Further, sequences that are not denoted as pool pick sequences, but that are included on the compact disks may also be used for gene silencing either alone or in combination with other sequences. However, their individual relative functionality would be less than that of a pool pick sequence. Table XIV, which is located in a file entitled Table_14.txt, provides an identification of genes by accession number, and Table XV, which is located in a file entitled Table_15.txt, provides a short name for the genes identified on the disk. The information contained on the disks is part of this patent application and are incorporated into the specification by reference. One may use these tables in order to identify functional siRNAs for the gene provided therein, by simply looking for the

gene of interest and an siRNA that is listed as functional. Preferably, one would select one or more of the siRNA that most optimized for the target of interest and is denoted as a pool pick.

5 **Table XII: siRNA selected by Formula VIII**

See data submitted herewith on a CD-ROM in accordance with PCT Administrative Instructions Section 801(a)

Table XIII: SMARTscores™

10 See data submitted herewith on a CD-ROM in accordance with PCT Administrative Instructions Section 801(a)

Table XIV: Identification of Targets

See data submitted herewith on a CD-ROM in accordance with PCT Administrative Instructions Section 801(a)

Table XV: Description of Targets

See data submitted herewith on a CD-ROM in accordance with PCT Administrative Instructions Section 801(a)

20

Many of the genes to which the described siRNA are directed play critical roles in disease etiology. For this reason, the siRNA listed in the accompanying compact disk may potentially act as therapeutic agents. A number of prophetic examples follow and should be understood in view of the siRNA that are identified on the accompanying CD. To isolate these siRNA, the appropriate message sequence for each gene is analyzed using one of the before mentioned formulas (preferably formula VIII) to identify potential siRNA targets. Subsequently these targets are BLAST'ed to eliminate homology with potentially off-targets.

30

The list of potential disease targets is extensive. For instance, over-expression of Bcl10 has been implicated in the development of MALT lymphoma (mucosa associated lymphoid tissue lymphoma) and thus, functional, highly functional, or hyperfunctional siRNA directed against that gene (*e.g.* SEQ. ID NO. 0427: GGAAACCUCUCAUUGCUGAA; SEQ. ID NO. 0428:

GAAAGAACCUUGCCGAUCA; SEQ. ID NO. 0429:

GGAAAUACAUCAGAGCUUA, or SEQ. ID NO. 0430:

GAAAGUAUGUGUCUUAAGU) may contribute to treatment of this disorder.

- 5 In another example, studies have shown that molecules that inhibit glutamine:fructose-6-phosphate aminotransferase (GFA) may act to limit the symptoms suffered by Type II diabetics. Thus, functional, highly functional, or hyperfunctional siRNA directed against GFA (also known as GFPT1: siRNA = SEQ. ID NO. 0433 UGAAACGGCUGCCUGAUUU; SEQ. ID NO. 0434
- 10 GAAGUUACCUCUUAUUAUU; SEQ. ID NO. 0435
GUACGAAACUGUAUGAUUA; SEQ. ID NO. 0436
GGACGAGGCUAUCAUUAUG) may contribute to treatment of this disorder.

- 15 In another example, the von Hippel-Lindau (VHL) tumor suppressor has been observed to be inactivated at a high frequency in sporadic clear cell renal cell carcinoma (RCC) and RCCs associated with VHL disease. The VHL tumor suppressor targets hypoxia-inducible factor-1 alpha (HIF-1 alpha), a transcription factor that can induce vascular endothelial growth factor (VEGF) expression, for ubiquitination and degradation. Inactivation of VHL can lead to increased levels of
- 20 HIF-1 alpha, and subsequent VEGF over expression. Such over expression of VEGF has been used to explain the increased (and possibly necessary) vascularity observed in RCC. Thus, functional, highly functional, or hyperfunctional siRNAs directed against either HIF-1 alpha (SEQ. ID NO. 0437 GAAGGAACCUUGAUGCUUUA; SEQ. ID NO. 0438 GCAUAUAUCUAGAAGGUAU; SEQ. ID NO. 0439
- 25 GAACAAAUACAUGGGAUUA; SEQ. ID NO. 0440
GGACACAGAUUUAGACUUG) or VEGF (SEQ. ID NO. 0441
GAACGUACUUGCAGAUUGUG; SEQ. ID NO. 0442
GAGAAAGCAUUUGUUUGUA; SEQ. ID NO. 0443
GGAGAAAGCAUUUGUUUGU; SEQ. ID NO. 0444
- 30 CGAGGCAGCUUGAGUUA) may be useful in the treatment of renal cell carcinoma.

In another example, gene expression of platelet derived growth factor A and B (PDGF-A and PDGF-B) has been observed to be increased 22- and 6-fold,

respectively, in renal tissues taken from patients with diabetic nephropathy as compared with controls. These findings suggest that over expression of PDGF A and B may play a role in the development of the progressive fibrosis that characterizes human diabetic kidney disease. Thus, functional, highly functional, or hyperfunctional

5 siRNAs directed against either PDGF A
(SEQ. ID NO. 0445: GGUAAGAUAUUGUGCUUUA;
SEQ. ID NO. 0446: CCGCAAAUAUGCAGAAUUA;
SEQ. ID NO. 0447: GGAUGUACAUGGCGUGUUA;
SEQ. ID NO. 0448: GGUGAAGUUUGUAUGUUUA) or

10

PDGF B

(SEQ. ID NO. 0449: CCGAGGAGCUUUAUGAGAU;
SEQ. ID NO. 0450: GCUCCGCGCUUCCGAUUU;
SEQ. ID NO. 0451 GAGCAGGAAUGGUGAGAUG;
15 SEQ. ID NO. 0452: GAACUUGGGAUAAGAGUGU;
SEQ. ID NO. 0453 CCGAGGAGCUUUAUGAGAU;
SEQ. ID NO. 0454 UUUAUGAGAUGCUGAGUGA) may be useful in the treatment of this form of kidney disorder.

20

In another example, a strong correlation exists between the over-expression of glucose transporters (*e.g.* GLUT12) and cancer cells. It is predicted that cells undergoing uncontrolled cell growth up-regulate GLUT molecules so that they can cope with the heightened energy needs associated with increased rates of proliferation and metastasis. Thus, siRNA-based therapies that target the molecules such as

25

GLUT1 (also known as SLC2A1: siRNA=
SEQ. ID NO.: 0455 GCAAUGAUGUCCAGAAGAA;
SEQ. ID NO.: 0456 GAAGAAUAUUCAGGACUUA;
SEQ. ID NO.: 0457 GAAGAGAGUCGGCAGAUGA;
SEQ. ID NO.: 0458 CCAAGAGUGUGCUAAAGAA)

30

GLUT12 (also known as SLCA12: siRNA =
SEQ. ID NO. 0459: GAGACACUCUGAAAUGAUA;
SEQ. ID NO. 0460: GAAAUGAUGUGGAUAAGAG;
SEQ. ID NO. 0461: GAUCAAAUCCUCCCUGAAA;

SEQ. ID NO. 0462: UGAAUGAGCUGAUGAUUGU) and other related transporters, may be of value in treating a multitude of malignancies.

5 The siRNA sequences listed above are presented in a 5' → 3' sense strand direction. In addition, siRNA directed against the targets listed above as well as those directed against other targets and listed in the accompanying compact disk may be useful as therapeutic agents.

Example VIII. Evidence for the Benefits of Pooling

10 Evidence for the benefits of pooling have been demonstrated using the reporter gene, luciferase. Ninety siRNA duplexes were synthesized using Dharmacon proprietary ACE® chemistry against one of the standard reporter genes: firefly luciferase. The duplexes were designed to start two base pairs apart and to cover approximately 180 base pairs of the luciferase gene (see sequences in **Table III**).
15 Subsequently, the siRNA duplexes were co-transfected with a luciferase expression reporter plasmid into HEK293 cells using standard transfection protocols and luciferase activity was assayed at 24 and 48 hours.

20 Transfection of individual siRNAs showed standard distribution of inhibitory effect. Some duplexes were active, while others were not. **Figure 15** represents a typical screen of ninety siRNA duplexes (SEQ. ID NO. 0032- 0120) positioned two base pairs apart. As the figure suggests, the functionality of the siRNA duplex is determined more by a particular sequence of the oligonucleotide than by the relative oligonucleotide position within a gene or excessively sensitive part of the mRNA,
25 which is important for traditional anti-sense technology.

When two continuous oligonucleotides were pooled together, a significant increase in gene silencing activity was observed. (See **Figure 16**) A gradual increase in efficacy and the frequency of pools functionality was observed when the number of
30 siRNAs increased to 3 and 4. (**Figures 16, 17**). Further, the relative positioning of the oligonucleotides within a pool did not determine whether a particular pool was functional (see **Figure 18**, in which 100% of pools of oligonucleotides distanced by 2, 10 and 20 base pairs were functional).

However, relative positioning may nonetheless have an impact. An increased functionality may exist when the siRNA are positioned continuously head to toe (5' end of one directly adjacent to the 3' end of the others).

- 5 Additionally, siRNA pools that were tested performed at least as well as the best oligonucleotide in the pool, under the experimental conditions whose results are depicted in **Figure 19**. Moreover, when previously identified non-functional and marginally (semi) functional siRNA duplexes were pooled together in groups of five at a time, a significant functional cooperative action was observed. (See **Figure 20**)
- 10 In fact, pools of semi-active oligonucleotides were 5 to 25 times more functional than the most potent oligonucleotide in the pool. Therefore, pooling several siRNA duplexes together does not interfere with the functionality of the most potent siRNAs within a pool, and pooling provides an unexpected significant increase in overall functionality

15

Example IX. Pooling Across Species

- Experiments were performed on the following genes: β -galactosidase, Renilla luciferase, and Secreted alkaline phosphatase, which demonstrates the benefits of pooling. (see **Figure 21**) Approximately 50% of individual siRNAs designed to
- 20 silence the above-specified genes were functional, while 100% of the pools that contain the same siRNA duplexes were functional.

Example X. Highly Functional siRNA

- Pools of five siRNAs in which each two siRNAs overlap to 10-90% resulted
- 25 in 98% functional entities (>80% silencing). Pools of siRNAs distributed throughout the mRNA that were evenly spaced, covering an approximate 20 – 2000 base pair range, were also functional. When the pools of siRNA were positioned continuously head to tail relative to mRNA sequences and mimicked the natural products of Dicer cleaved long double stranded RNA, 98% of the pools evidenced highly functional
- 30 activity (>95% silencing).

Example XI. Human cyclophyline

Table III above lists the siRNA sequences for the human cyclophyline protein. A particularly functional siRNA may be selected by applying these sequences to any of Formula I to VII above.

Alternatively, one could pool 2, 3, 4, 5 or more of these sequences to create a kit for silencing a gene. Preferably, within the kit there would be at least one sequence that has a relatively high predicted functionality when any of Formulas I - VII is applied.

10 Example XII. Sample Pools of siRNAs and Their Application to Human Disease

The genetic basis behind human disease is well documented and siRNA may be used as both research or diagnostic tools and therapeutic agents, either individually or in pools. Genes involved in signal transduction, the immune response, apoptosis, DNA repair, cell cycle control, and a variety of other physiological functions have clinical relevance and therapeutic agents that can modulate expression of these genes may alleviate some or all of the associated symptoms. In some instances, these genes can be described as a member of a family or class of genes and siRNA (randomly, conventionally, or rationally designed) can be directed against one or multiple members of the family to induce a desired result.

20 To identify rationally designed siRNA to each gene, the sequence was analyzed using Formula VIII to identify a SMARTpool containing the functional sequences. To confirm the activity of these sequences, the siRNA are introduced into a cell type of choice (*e.g.* HeLa cells, HEK293 cells) and the levels of the appropriate
25 message are analyzed using one of several art proven techniques. SiRNA having heightened levels of potency can be identified by testing each of the before mentioned duplexes at increasingly limiting concentrations. Similarly, siRNA having increased levels of longevity can be identified by introducing each duplex into cells and testing functionality at 24, 48, 72, 96, 120, 144, 168, and 192 hours after transfection. Agents
30 that induce >95% silencing at sub-nanomolar concentrations and/or induce functional levels of silencing for >96 hours are considered hyperfunctional.

The following are non-limiting examples of families of proteins to which siRNA described in this document are targeted against:

Transporters, Pumps, and Channels

5 Transporters, pumps, and channels represent one class of genes that are attractive targets for siRNAs. One major class of transporter molecules are the ATP-binding cassette (ABC) transporters. To date, nearly 50 human ABC-transporter genes have been characterized and have been shown to be involved in a variety of physiological functions including transport of bile salts, nucleosides, chloride ions, 10 cholesterol, toxins, and more. Predominant among this group are MDR1 (which encodes the P-glycoprotein, NP_000918), the MDR-related proteins (MRP1-7), and the breast cancer resistance protein (BCRP). In general, these transporters share a common structure, with each protein containing a pair of ATP-binding domains (also known as nucleotide binding folds, NBF) and two sets of transmembrane (TM) 15 domains, each of which typically contains six membrane-spanning α -helices. The genes encoding this class of transporter are organized as either full transporters (*i.e.* containing two TM and two NBF domains) or as half transporters that assemble as either homodimers or heterodimers to create functional transporters. As a whole, members of the family are widely dispersed throughout the genome and show a high 20 degree of amino acid sequence identity among eukaryotes.

ABC-transporters have been implicated in several human diseases. For instance, molecular efflux pumps of this type play a major role in the development of drug resistance exhibited by a variety of cancers and pathogenic microorganisms. In 25 the case of human cancers, increased expression of the MDR1 gene and related pumps have been observed to generate drug resistance to a broad collection of commonly used chemotherapeutics including doxorubicin, daunorubicin, vinblastine, vincristine, colchicines. In addition to the contribution these transporters make to the development of multi-drug resistance, there are currently 13 human genetic diseases 30 associated with defects in 14 different transporters. The most common of these conditions include cystic fibrosis, Stargardt disease, age-related macular degeneration, adrenoleukodystrophy, Tangier disease, Dubin-Johnson syndrome and progressive familial intrahepatic cholestasis. For this reason, siRNAs directed against members of this, and related, families are potentially valuable research and therapeutic tools.

With respect to channels, analysis of *Drosophila* mutants has enabled the initial molecular isolation and characterization of several distinct channels including (but not limited to) potassium (K^+) channels. This list includes shaker (Sh), which encodes a voltage activated K^+ channel, slowpoke (Slo), a Ca^{2+} activated K^+ channel, and ether-a-go-go (Eag). The Eag family is further divided into three subfamilies: Eag, Elk (eag-like K channels), and Erg (Eag related genes).

The Erg subfamily contains three separate family members (Erg1-3) that are distantly related to the sh family of voltage activated K^+ channels. Like sh, erg polypeptides contain the classic six membrane spanning architecture of K^+ channels (S1-S6) but differ in that each includes a segment associated with the C-terminal cytoplasmic region that is homologous to cyclic nucleotide binding domains (cNBD). Like many isolated ion channel mutants, erg mutants are temperature-sensitive paralytics, a phenotype caused by spontaneous repetitive firing (hyperactivity) in neurons and enhanced transmitter release at the neuromuscular junction.

Initial studies on the tissue distribution of all three members of the erg subfamily show two general patterns of expression. Erg1 and erg3 are broadly expressed throughout the nervous system and are observed in the heart, the superior mesenteric ganglia, the celiac ganglia, the retina, and the brain. In contrast, erg2 shows a much more restricted pattern of expression and is only observed in celiac ganglia and superior mesenteric ganglia. Similarly, the kinetic properties of the three erg potassium channels are not homogeneous. Erg1 and erg2 channels are relatively slow activating delayed rectifiers whereas the erg3 current activates rapidly and then exhibits a predominantly transient component that decays to a sustained plateau. The current properties of all three channels are sensitive to methanesulfonanilides, suggesting a high degree of conservation in the pore structure of all three proteins.

Recently, the erg family of K^+ channels has been implicated in human disease. Consistent with the observation that erg1 is expressed in the heart, single strand conformation polymorphism and DNA sequence analyses have identified HERG (human erg1) mutations in six long-QT-syndrome (LQT) families, an inherited disorder that results in sudden death from a ventricular tachyarrhythmia. Thus siRNA

directed against this group of molecules (*e.g.* KCNH1-8) will be of extreme therapeutic value.

Another group of channels that are potential targets of siRNAs are the CLCA family that mediate a Ca^{2+} -activated Cl^- conductance in a variety of tissues. To date, two bovine (bCLC1; bCLCA2 (Lu-ECAM-1)), three mouse (mCLCA1; mCLCA2; mCLCA3) and four human (hCLCA1; hCLCA2; hCLCA3; hCLCA4) CLCA family members have been isolated and patch-clamp studies with transfected human embryonic kidney (HEK-293) cells have shown that bCLCA1, mCLCA1, and hCLCA1 mediate a Ca^{2+} -activated Cl^- conductance that can be inhibited by the anion channel blocker DIDS and the reducing agent dithiothreitol (DTT).

The protein size, structure, and processing seem to be similar among different CLCA family members and has been studied in greatest detail for Lu-ECAM-1. The Lu-ECAM-1 open reading frame encodes a precursor glycoprotein of 130 kDa that is processed to a 90-kDa amino-terminal cleavage product and a group of 30- to 40-kDa glycoproteins that are glycosylation variants of a single polypeptide derived from its carboxy terminus. Both subunits are associated with the outer cell surface, but only the 90-kDa subunit is thought to be anchored to the cell membrane via four transmembrane domains.

Although the protein processing and function appear to be conserved among CLCA homologs, significant differences exist in their tissue expression patterns. For example, bovine Lu-ECAM-1 is expressed primarily in vascular endothelia, bCLCA1 is exclusively detected in the trachea, and hCLCA1 is selectively expressed in a subset of human intestinal epithelial cells. Thus the emerging picture is that of a multigene family with members that are highly tissue specific, similar to the ClC family of voltage-gated Cl^- channels. The human channel, hCLCA2, is particularly interesting from a medical and pharmacological standpoint. CLCA2 is expressed on the luminal surface of lung vascular endothelia and serves as an adhesion molecule for lung metastatic cancer cells, thus mediating vascular arrest and lung colonization. Expression of this molecule in normal mammary epithelium is consistently lost in human breast cancer and in nearly all tumorigenic breast cancer cell lines. Moreover,

re-expression of hCLCA2 in human breast cancer cells abrogates tumorigenicity in nude mice, implying that hCLCA2 acts as a tumour suppressor in breast cancer. For these reasons, siRNA directed against CLCA family members and related channels may prove to be valuable in research and therapeutic venues.

5

Transporters Involved in Synaptic Transmission.

Synaptic transmission involves the release of a neurotransmitter into the synaptic cleft, interaction of that transmitter with a postsynaptic receptor, and subsequent removal of the transmitter from the cleft. In most synapses the signal is terminated by a rapid reaccumulation of the neurotransmitter into presynaptic terminals. This process is catalyzed by specific neurotransmitter transporters that are often energized by the electrochemical gradient of sodium across the plasma membrane of the presynaptic cells.

10

Aminobutyric acid (GABA) is the major inhibitory neurotransmitter in the central nervous system. The inhibitory action of GABA, mediated through GABA_A/GABA_B receptors, and is regulated by GABA transporters (GATs), integral membrane proteins located perisynaptically on neurons and glia. So far four different carriers (GAT1-GAT4) have been cloned and their cellular distribution has been partly worked out. Comparative sequence analysis has revealed that GABA transporters are related to several other proteins involved in neurotransmitter uptake including gamma-aminobutyric acid transporters, monoamine transporters, amino acid transporters, certain "orphan" transporters, and the recently discovered bacterial transporters. Each of these proteins has a similar 12 transmembrane helices topology and relies upon the Na⁺/Cl⁻ gradient for transport function. Transport rates are dependent on substrate concentrations, with half-maximal effective concentrations for transport frequently occurring in the submicromolar to low micromolar range. In addition, transporter function is bidirectional, and non-vesicular efflux of transmitter may contribute to ambient extracellular transmitter levels.

20

25

30

Recent evidence suggests that GABA transporters, and neurotransmitter transporters in general, are not passive players in regulating neuronal signaling; rather, transporter function can be altered by a variety of initiating factors and signal transduction cascades. In general, this functional regulation occurs in two ways,

either by changing the rate of transmitter flux through the transporter or by changing the number of functional transporters on the plasma membrane. A recurring theme in transporter regulation is the rapid redistribution of the transporter protein between intracellular locations and the cell surface. In general, this functional modulation occurs in part through activation of second messengers such as kinases, phosphatases, arachidonic acid, and pH. However, the mechanisms underlying transporter phosphorylation and transporter redistribution have yet to be fully elucidated.

GABA transporters play a pathophysiological role in a number of human diseases including temporal lobe epilepsy and are the targets of pharmacological interventions. Studies in seizure sensitive animals show some (but not all) of the GAT transporters have altered levels of expression at times prior to and post seizure, suggesting this class of transporter may affect epileptogenesis, and that alterations following seizure may be compensatory responses to modulate seizure activity. For these reasons, siRNAs directed against members of this family of genes (including but not limited to SLC6A1-12) may prove to be valuable research and therapeutic tools.

Organic Ion Transporters.

The human body is continuously exposed to a great variety of xenobiotics, via food, drugs, occupation, and environment. Excretory organs such as kidney, liver, and intestine defend the body against the potentially harmful effects of these compounds by transforming them into less active metabolites that are subsequently secreted from the system.

Carrier-mediated transport of xenobiotics and their metabolites exist for the active secretion of organic anions and cations. Both systems are characterized by a high clearance capacity and tremendous diversity of substances accepted, properties that result from the existence of multiple transporters with overlapping substrate specificities. The class of organic anion transporters plays a critical role in the elimination of a large number of drugs (*e.g.*, antibiotics, chemotherapeutics, diuretics, nonsteroidal anti-inflammatory drugs, radiocontrast agents, cytostatics); drug metabolites (especially conjugation products with glutathione, glucuronide, glycine, sulfate, acetate); and toxicants and their metabolites (*e.g.*, mycotoxins, herbicides,

plasticizers, glutathione *S*-conjugates of polyhaloalkanes, polyhaloalkenes, hydroquinones, aminophenols), many of which are specifically harmful to the kidney.

Over the past couple of years the number of identified anion transporting molecules has grown tremendously. Uptake of organic anions (OA^-) across the basolateral membrane is mediated by the classic sodium-dependent organic anion transport system, which includes α -ketoglutarate ($\alpha\text{-KG}^{2-}$)/ OA^- exchange via the organic anion transporter (OAT1) and sodium-ketoglutarate cotransport via the Na^+ /dicarboxylate cotransporter (SDCT2). The organic anion transporting polypeptide, Oatp1, and the kidney-specific OAT-K1 and OAT-K2 are seen as potential molecules that mediate facilitated OA^- efflux but could also be involved in reabsorption via an exchange mechanism. Lastly the PEPT1 and PEPT2 mediate luminal uptake of peptide drugs, whereas CNT1 and CNT2 are involved in reabsorption of nucleosides

The organic anion-transporting polypeptide 1 (Oatp1) is a Na^+ - and ATP-independent transporter originally cloned from rat liver. The tissue distribution and transport properties of the Oatp1 gene product are complex. Oatp1 is localized to the basolateral membrane of hepatocytes, and is found on the apical membrane of S3 proximal tubules. Studies with transiently transfected cells (*e.g.* HeLa cells) have indicated that Oatp1 mediates transport of a variety of molecules including taurocholate, estrone-3-sulfate, aldosterone, cortisol, and others. The observed uptake of taurocholate by Oatp1 expressed in *X. laevis* oocytes is accompanied by efflux of GSH, suggesting that transport by this molecule may be glutathione dependent.

Computer modeling suggests that members of the Oatp family are highly conserved, hydrophobic, and have 12 transmembrane domains. Decreases in expression of Oatp family members have been associated with cholestatic liver diseases and human hepatoblastomas, making this family of proteins of key interest to researchers and the medical community. For these reasons, siRNAs directed against OAT family members (including but not limited to SLC21A2, 3, 6, 8, 9, 11, 12, 14, 15, and related transporters) are potentially useful as research and therapeutic tools.

Nucleoside transporters.

Nucleoside transporters play key roles in physiology and pharmacology.

Uptake of exogenous nucleosides is a critical first step of nucleotide synthesis in tissues such as bone marrow and intestinal epithelium and certain parasitic organisms that lack *de novo* pathways for purine biosynthesis. Nucleoside transporters also control the extracellular concentration of adenosine in the vicinity of its cell surface receptors and regulate processes such as neurotransmission and cardiovascular activity. Adenosine itself is used clinically to treat cardiac arrhythmias, and nucleoside transport inhibitors such as dipyridamole, dilazep, and draflazine function as coronary vasodilators.

In mammals, plasma membrane transport of nucleosides is brought about by members of the concentrative, Na^+ -dependent (CNT) and equilibrative, Na^+ -independent (ENT) nucleoside transporter families. CNTs are expressed in a tissue-specific fashion; ENTs are present in most, possibly all, cell types and are responsible for the movement of hydrophilic nucleosides and nucleoside analogs down their concentration gradients. In addition, structure/function studies of ENT family members have predicted these molecules to contain eleven transmembrane helical segments with an amino terminus that is intracellular and a carboxyl terminus that is extracellular. The proteins have a large glycosylated loop between TMs 1 and 2 and a large cytoplasmic loop between TMs 6 and 7. Recent investigations have implicated the TM 3-6 region as playing a central role in solute recognition. The medical importance of the ENT family of proteins is broad. In humans adenosine exerts a range of cardioprotective effects and inhibitors of ENTs are seen as being valuable in alleviating a variety of cardio/cardiovascular ailments. In addition, responses to nucleoside analog drugs has been observed to vary considerably amongst *e.g.* cancer patients. While some forms of drug resistance have been shown to be tied to the up-regulation of ABC-transporters (*e.g.* MDR1), resistance may also be the result of reduced drug uptake (*i.e.* reduced ENT expression). Thus, a clearer understanding of ENT transporters may aid in optimizing drug treatments for patients suffering a wide range of malignancies. For these reasons, siRNAs directed against this class of molecules (including SLC28A1-3, SLC29A1-4, and related molecules) may be useful as therapeutic and research tools.

Sulfate Transporters.

All cells require inorganic sulfate for normal function. Sulfate is the fourth most abundant anion in human plasma and is the major source of sulfur in many organisms. Sulfation of extracellular matrix proteins is critical for maintaining normal cartilage metabolism and sulfate is an important constituent of myelin membranes found in the brain

Because sulfate is a hydrophilic anion that cannot passively cross the lipid bilayer of cell membranes, all cells require a mechanism for sulfate influx and efflux to ensure an optimal supply. To date, a variety of sulfate transporters have been identified in tissues from many origins. These include the renal sulfate transporters (NaSi-1 and Sat-1), the ubiquitously expressed diastrophic dysplasia sulfate transporter (DTDST), the intestinal sulfate transporter (DRA), and the erythrocyte anion exchanger (AE1). Most, if not all, of these molecules contain the classic 12 transmembrane spanning domain architecture commonly found amongst members of the anion transporter superfamily.

Recently three different sulfate transporters have been associated with specific human genetic diseases. Family members SLC26A2, SLC26A3, and SLC26A4 have been recognized as the disease genes mutated in diastrophic dysplasia, congenital chloride diarrhea (CLD), and Pendred syndrome (PDS), respectively. DTDST is a particularly complex disorder. The gene encoding this molecule maps to chromosome 5q, and encodes two distinct transcripts due to alternative exon usage. In contrast to other sulfate transporters (*e.g.* Sat-1) anion movement by the DTDST protein is markedly inhibited by either extracellular chloride or bicarbonate. Impaired function of the DTDST gene product leads to undersulfation of proteoglycans and a complex family of recessively inherited osteochondrodysplasias (achondrogenesis type 1B, atelosteogenesis type II, and diastrophic dysplasia) with clinical features including but not limited to, dwarfism, spinal deformation, and specific joint abnormalities. Interestingly, while epidemiological studies have shown that the disease occurs in most populations, it is particularly prevalent in Finland owing to an apparent founder effect. For these reasons, siRNAs directed against this class of genes (including but not limited to SLC26A1-9, and related molecules) may be potentially helpful in both therapeutic and research venues.

Ion Exchangers

Intracellular pH regulatory mechanisms are critical for the maintenance of countless cellular processes. For instance, in muscle cells, contractile processes and metabolic reactions are influenced by pH. During periods of increased energy demands and ischemia, muscle cells produce large amounts of lactic acid that, without quick and efficient disposal, would lead to acidification of the sarcoplasm.

Several different transport mechanisms have evolved to maintain a relatively constant intracellular pH. The relative contribution of each of these processes varies with cell type, the metabolic requirements of the cell, and the local environmental conditions. Intracellular pH regulatory processes that have been characterized functionally include but are not limited to the Na^+/H^+ exchange, the $\text{Na}(\text{HCO}_3)_n$ cotransport, and the Na^+ -dependent and -independent Cl^-/base exchangers. As bicarbonate and CO_2 comprise the major pH buffer of biological fluids, sodium biocarbonate cotransporters (NBCs) are critical. Studies have shown that these molecules exist in numerous tissues including the kidney, brain, liver, cornea, heart, and lung, suggesting that NBCs play an important role in mediating HCO_3^- transport in both epithelial as well as nonepithelial cells.

Recent molecular cloning experiments have identified the existence of four NBC isoforms (NBC1, 2, 3 and 4) and two NBC-related proteins, AE4 and NCBE (Anion Exchanger 4 and Na-dependent Chloride-Bicarbonate Exchanger). The secondary structure analyses and hydropathy profile of this family predict them to be intrinsic membrane proteins with 12 putative transmembrane domains and several family members exhibit *N*-linked glycosylation sites, protein kinases A and C, casein kinase II, and ATP/GTP-binding consensus phosphorylation sites, as well as potential sites for myristylation and amidation. AE4 is a relatively recent addition to this family of proteins and shows between 30-48% homology with the other family members. When expressed in COS-7 cells and *Xenopus* oocytes AE4 exhibits sodium-independent and DIDS-insensitive anion exchanger activity. Exchangers have been shown to be responsible for a variety of human diseases. For instance, mutations in three genes of the anion transporter family (SLC) are believed to cause known hereditary diseases, including chondrodysplasia (SLC26A2, DTD), diarrhea (A3, down-regulated in adenoma/chloride-losing diarrhea protein: DRA/CLD), and

goiter/deafness syndrome (A4, pendrin). Moreover, mutations in $\text{Na}^+/\text{HCO}_3^-$ co-transporters have also been associated with various human maladies. For these reasons, siRNAs directed against these sorts of genes (*e.g.* SLC4A4-10, and related genes) may be useful for therapeutic and research purposes.

5

Receptors Involved in Synaptic Transmission

In all vertebrates, fast inhibitory synaptic transmission is the result of the interaction between the neurotransmitters glycine (Gly) and γ -aminobutyric acid (GABA) and their respective receptors. The strychnine-sensitive glycine receptor is especially important in that it acts in the mammalian spinal cord and brain stem and has a well-established role in the regulation of locomotor behavior.

10

Glycine receptors display significant sequence homology to several other receptors including the nicotinic acetylcholine receptor, the aminobutyric acid receptor type A (GABA_AR), and the serotonin receptor type 3 (5-HT₃R) subunits. As members of the superfamily of ligand-gated ion channels, these polypeptides share common topological features. The glycine receptor is composed of two types of glycosylated integral membrane proteins (α 1- α 4 and β) arranged in a pentameric suprastructure. The alpha subunit encodes a large extracellular, N-terminal domain that carries the structural determinants essential for agonist and antagonist-binding, followed by four transmembrane spanning regions (TM1-TM4), with TM2 playing the critical role of forming the inner wall of the chloride channel.

15

20

The density, location, and subunit composition of glycine neurotransmitter receptors changes over the course of development. It has been observed that the amount of GlyR gene translation (assessed by the injection of developing rat cerebral cortex mRNA into *Xenopus* oocytes) decreases with age, whereas that of GABA_ARs increases. In addition, the type and location of mRNAs coding for GlyR changes over the course of development. For instance in a study of the expression of alpha 1 and alpha 2 subunits in the rat, it was observed that (in embryonic periods E11-18) the mantle zone was scarce in the alpha 1 mRNA, but the germinal zone (matrix layer) at E11-14 expressed higher levels of the message. At postnatal day 0 (P0), the alpha 1 signals became manifested throughout the gray matter of the spinal cord. By contrast,

30

the spinal tissues at P0 exhibited the highest levels of alpha 2 mRNA, which decreased with the postnatal development.

In both, man and mouse mutant lines, mutations of GlyR subunit genes result in hereditary motor disorders characterized by exaggerated startle responses and increased muscle tone. Pathological alleles of the *Glr1* gene are associated with the murine phenotypes *oscillator* (*spd^{ot}*) and *spasmodic* (*spd*). Similarly, a mutant allele of *Glr2* has been found to underly the molecular pathology of the *spastic* mouse (*spa*). Resembling the situation in the mouse, a variety of *GLRA1* mutant alleles have been shown to be associated with the human neurological disorder hyperekplexia or startle disease. For these reasons, siRNA directed against glycine receptors (GLRA1-3, GLRB, and related molecules), glutamate receptors, GABA receptors, ATP receptors, and related neurotransmitter receptor molecules may be valuable therapeutic and research reagents.

Proteases

Kallikreins

One important class of proteases are the kallikreins, serine endopeptidases that split peptide substrates preferentially on the C-terminal side of internal arginyl and lysyl residues. Kallikreins are generally divided into two distinct groups, plasma kallikreins and tissue kallikreins. Tissue kallikreins represent a large group of enzymes that have substantial similarities at both the gene and protein level. The genes encoding this group are frequently found on a single chromosome, are organized in clusters, and are expressed in a broad range of tissues (*e.g.* pancreas, ovaries, breast). In contrast, the plasma form of the enzyme is encoded by a single gene (*e.g.* KLK3) that has been localized to chromosome 4q34-35 in humans. The gene encoding plasma kallikrein is expressed solely in the liver, contains 15 exons, and encodes a glycoprotein that is translated as a preprotein called prekallikrein.

Kallikreins are believed to play an important role in a host of physiological events. For instance, the immediate consequence of plasma prekallikrein activation is the cleavage of high molecular weight kininogen (HK) and the subsequent liberation of bradykinin, a nine amino acid vasoactive peptide that is an important mediator of

inflammatory responses. Similarly, plasma kallikrein promotes single-chain urokinase activation and subsequent plasminogen activation, events that are critical to blood coagulation and wound healing.

- 5 Disruptions in the function of kallikreins have been implicated in a variety of pathological processes including imbalances in renal function and inflammatory processes. For these reasons, siRNAs directed against this class of genes (*e.g.* KLK1-15) may prove valuable in both research and therapeutic settings.

10 ADAM Proteins

- The process of fertilization takes place in a series of discrete steps whereby the sperm interacts with; i) the cumulus cells and the hyaluronic acid extracellular matrix (ECM) in which they are embedded, ii) the egg's own ECM, called the *zona pellucida* (ZP), and iii) the egg plasma membrane. During the course of these interactions, the
- 15 "acrosome reaction," the exocytosis of the acrosome vesicle on the head of the sperm, is induced, allowing the sperm to penetrate the ZP and gain access to the perivitelline space. This process exposes new portions of the sperm membrane, including the inner acrosomal membrane and the equatorial segment, regions of the sperm head that can participate in initial gamete membrane binding.

20

- The interactions of the gamete plasma membranes appear to involve multiple ligands and receptors and are frequently compared to leukocyte-endothelial interactions. These interactions lead to a series of signal transduction events in the egg, known as collectively as egg activation and include the initiation of oscillations
- 25 in intracellular calcium concentration, the exit from meiosis, the entry into the first embryonic mitosis, and the formation of a block to polyspermy via the release of ZP-modifying enzymes from the egg's cortical granules. Ultimately, sperm and egg not only adhere to each other but also go on to undergo membrane fusion, making one cell (the zygote) from two.

30

 Studies on the process of sperm-egg interactions have identified a number of proteins that are crucial for fertilization. One class of proteins, called the ADAM family (A Disintegrin And Metalloprotease), has been found to be important in spermatogenesis and fertilization, as well as various developmental systems including

myogenesis and neurogenesis. Members of the family contain a disintegrin and metalloprotease domain (and therefore have (potentially) both cell adhesion and protease activities), as well as cysteine-rich regions, epidermal growth factor (EGF)-like domains, a transmembrane region, and a cytoplasmic tail. Currently, the ADAM gene family has 29 members and constituents are widely distributed in many tissues including the brain, testis, epididymis, ovary, breast, placenta, liver, heart, lung, bone, and muscle.

One of the best-studied members of the ADAM family is fertilin, a heterodimeric protein comprised of at least two subunits, fertilin alpha and fertilin beta. The fertilin beta gene (ADAM2) has been disrupted with a targeting gene construct corresponding to the exon encoding the fertilin beta disintegrin domain. Sperm from males homozygous for disruptions in this region exhibit defects in multiple facets of sperm function including reduced levels of sperm transit from the uterus to the oviduct, reduced sperm-ZP binding, and reduced sperm-egg binding, all of which contribute to male infertility.

Recently, four new ADAM family members (ADAM 24-27) have been isolated. The deduced amino acid sequences show that all four contain the complete domain organization common to ADAM family members and Northern Blot analysis has shown all four to be specific to the testes. SiRNAs directed against this class of genes (*e.g.* ADAM2 and related proteins) may be useful as research tools and therapeutics directed toward fertility and birth control.

25 Aminopeptidases

Aminopeptidases are proteases that play critical roles in processes such as protein maturation, protein digestion in its terminal stage, regulation of hormone levels, selective or homeostatic protein turnover, and plasmid stabilization. These enzymes generally have broad substrate specificity, occur in several forms and play a major role in physiological homeostasis. For instance, the effects of bradykinin, angiotensin converting enzyme (ACE), and other vasoactive molecules are muted by one of several peptidases that cleave the molecule at an internal position and eliminate its ability to bind its cognate receptor (*e.g.* for bradykinin, the B2-receptor).

Among the enzymes that can cleave bradykinin is the membrane bound aminopeptidase P, also referred to as aminoacylproline aminopeptidase, proline aminopeptidase; X-Pro aminopeptidase (eukaryote) and XPNPEP2. Aminopeptidase P is an aminoacylproline aminopeptidase specific for NH₂-terminal Xaa-proline bonds. The enzyme i) is a mono-zinc-containing molecule that lacks any of the typical metal binding motifs found in other zinc metalloproteases, ii) has an active-site configuration similar to that of other members of the MG peptidase family, and iii) is present in a variety of tissues including but not limited to the lung, kidney, brain, and intestine.

10

Aminopeptidases play an important role in a diverse set of human diseases. Low plasma concentrations of aminopeptidase P are a potential predisposing factor for development of angio-oedema in patients treated with ACE inhibitors, and inhibitors of aminopeptidase P may act as cardioprotectors against other forms of illness including, but not limited to myocardial infarction. For these reasons, siRNAs directed against this family of proteins (including but not limited to XPNPEP1 and related proteins) may be useful as research and therapeutic tools.

15

Serine Proteases

One important class of proteases are the serine proteases. Serine proteases share a common catalytic triad of three amino acids in their active site (serine (nucleophile), aspartate (electrophile), and histidine (base)) and can hydrolyze either esters or peptide bonds utilizing mechanisms of covalent catalysis and preferential binding of the transition state. Based on the position of their introns serine proteases have been classified into a minimum of four groups including those in which 1) the gene has no introns interrupting the exon coding for the catalytic triad (*e.g.* the haptoglobin gene,); 2) each gene contains an intron just downstream from the codon for the histidine residue at the active site, a second intron downstream from the exon containing the aspartic acid residue of the active site and a third intron just upstream from the exon containing the serine of the active site (*e.g.* trypsinogen, chymotrypsinogen, kallikrein and proelastase); 3) the genes contain seven introns interrupting the exons coding the catalytic region (*e.g.* complement factor B gene); and 4) the genes contain two introns resulting in a large exon that contains both the

25

30

active site aspartic acid and serine residues (*e.g.* factor X, factor IX and protein C genes).

Cytotoxic lymphocytes (*e.g.* CD8(+) cytotoxic T cells and natural killer cells) form the major defense of higher organisms against virus-infected and transformed cells. A key function of these cells is to detect and eliminate potentially harmful cells by inducing them to undergo apoptosis. This is achieved through two principal pathways, both of which require direct but transient contact between the killer cell and its target. The first pathway involves ligation of TNF receptor-like molecules such as Fas/CD95 to their cognate ligands, and results in mobilization of conventional, programmed cell-death pathways centered on activation of pro-apoptotic caspases. The second mechanism consists of a pathway whereby the toxic contents of a specialized class of secretory vesicles are introduced into the target cell. Studies over the last two decades have identified the toxic components as Granzymes, a family of serine proteases that are expressed exclusively by cytotoxic T lymphocytes and natural killer (NK) cells. These agents are stored in specialized lytic granules and enter the target cell via endocytosis. Like caspases, cysteine proteases that play an important role in apoptosis, granzymes can cleave proteins after acidic residues, especially aspartic acid, and induce apoptosis in the recipient cell.

Granzymes have been grouped into three subfamilies according to substrate specificity. Members of the granzyme family that have enzymatic activity similar to the serine protease chymotrypsin are encoded by a gene cluster termed the 'chymase locus'. Similarly, granzymes with trypsin-like specificities are encoded by the 'tryptase locus', and a third subfamily cleaves after unbranched hydrophobic residues, especially methionine, and are encoded by the 'Met-ase locus'. All granzymes are synthesized as zymogens and, after clipping of the leader peptide, obtain maximal enzymatic activity subsequent to the removal of an amino-terminal dipeptide.

Granzymes have been found to be important in a number of important biological functions including defense against intracellular pathogens, graft versus host reactions, the susceptibility to transplantable and spontaneous malignancies, lymphoid homeostasis, and the tendency toward auto-immune diseases. For these

reasons, siRNAs directed against granzymes (e.g. GZMA, GZMB, GZMH, GZHK, GZMM) and related serine proteases may be useful research and therapeutic reagents.

Kinases

5 Protein Kinases (PKs) have been implicated in a number of biological processes. Kinase molecules play a central role in modulating cellular physiology and developmental decisions, and have been implicated in a large list of human maladies including cancer, diabetes, and others.

10 During the course of the last three decades, over a hundred distinct protein kinases have been identified, all with presumed specific cellular functions. A few of these enzymes have been isolated to sufficient purity to perform *in vitro* studies, but most remain intractable due to the low abundance of these molecules in the cell. To counter this technical difficulty, a number of protein kinases have been isolated by
15 molecular cloning strategies that utilize the conserved sequences of the catalytic domain to isolate closely related homologs. Alternatively, some kinases have been purified (and subsequently studied) based on their interactions with other molecules.

p58 is a member of the p34cdc2-related supergene family and contains a large
20 domain that is highly homologous to the cell division control kinase, cdc2. This new cell division control-related protein kinase was originally identified as a component of semipurified galactosyltransferase; thus, it has been denoted galactosyltransferase-associated protein kinase (GTA-kinase). GTA-kinase has been found to be expressed in both adult and embryonic tissues and is known to phosphorylate a number of
25 substrates, including histone H1, and casein. Interestingly enough, over expression of this molecule in CHO cells has shown that elevated levels of p58 result in a prolonged late telophase and an early G1 phase, thus hinting of an important role for GTA-kinase in cell cycle regulation.

30 Cyclin Dependent Kinases

The cyclin-dependent kinases (Cdks) are a family of highly conserved serine/threonine kinases that mediate many of the cell cycle transitions that occur during duplication. Each of these Cdk catalytic subunits associates with a specific

subset of regulatory subunits, termed cyclins, to produce a distinct Cdk-cyclin kinase complex that, in general, functions to execute a unique cell cycle event.

Activation of the Cdk-cyclin kinases during cellular transitions is controlled by a variety of regulatory mechanisms. For the Cdc2-cyclin B complex, inhibition of kinase activity during S phase and G₂ is accomplished by phosphorylation of two Cdc2 residues, Thr¹⁴ and Tyr¹⁵, which are positioned within the ATP-binding cleft. Phosphorylation of Thr¹⁴ and/or Tyr¹⁵ suppresses the catalytic activity of the molecule by disrupting the orientation of the ATP present within this cleft. In contrast, the abrupt dephosphorylation of these residues by the Cdc25 phosphatase results in the rapid activation of Cdc2-cyclin B kinase activity and subsequent downstream mitotic events. While the exact details of this pathway have yet to be elucidated, it has been proposed that Thr¹⁴/Tyr¹⁵ phosphorylation functions to permit a cell to attain a critical concentration of inactive Cdk-cyclin complexes, which, upon activation, induces a rapid and complete cell cycle transition. Furthermore, there is evidence in mammalian cells that Thr¹⁴/Tyr¹⁵ phosphorylation also functions to delay Cdk activation after DNA damage.

The *Schizosaccharomyces pombe wee1* gene product was the first kinase identified that is capable of phosphorylating Tyr¹⁵ in Cdc2. Homologs of the Wee1 kinase have been subsequently identified and biochemically characterized from a wide range of species including human, mouse, frog, *Saccharomyces cerevisiae*, and *Drosophila*. In vertebrate systems, where Thr¹⁴ in Cdc2 is also phosphorylated, the Wee1 kinase was capable of phosphorylating Cdc2 on Tyr¹⁵, but not Thr¹⁴, indicating that another kinase was responsible for Thr¹⁴ phosphorylation. This gene, Myt1 kinase, was recently isolated from the membrane fractions of *Xenopus* egg extracts and has been shown to be capable of phosphorylating Thr¹⁴ and, to a lesser extent, Tyr¹⁵ in Cdc2. A human Myt1 homolog displaying similar properties has been isolated, as well as a non-membrane-associated molecule with Thr¹⁴ kinase activity.

In the past decade it has been shown that cancer can originate from overexpression of positive regulators, such as cyclins, or from underexpression of negative regulators (e.g. p16 (INK4a), p15 (INK4b), p21 (Cip1)). Inhibitors such as Myt1 are the focus of much cancer research because they are capable of controlling

cell cycle proliferation, now considered the Holy Grail for cancer treatment. For these reasons, siRNA directed against kinases and kinase inhibitors including but not limited to ABL1, ABL2, ACK1, ALK, AXL, BLK, BMX, BTK, C20orf64, CSF1R, SCK, DDR1, DDR2, DKFZp761P1010, EGFR, EPHA1, EPHA2, EPHA3, EPHA4, 5 EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, EPHB6, ERBB2, ERBB3, ERBB4, FER, FES, FGFR1, FGFR2, FGFR3, FGFR4, FGR, FLT1, FLT3, FLT4, FRK, FYN, HCK, IGF1R, INSR, ITK, JAK1, JAK2, JAK3, KDR, KIAA1079, KIT, LCK, LTK, LYN, MATK, MERTK, MET, MST1R, MUSK, NTRK1, NTRK2, NTRK3, PDGFRA, PDGFRB, PTK2, PTK2B, PTK6, PTK7, PTK9, PTK9L, RET, 10 ROR1, ROR2, ROS1, RYK, SRC, SYK, TEC, TEK, TIE, TNK1, TXK, TYK2, TYRO3, YES1, and related proteins, may be useful for research and therapeutic purposes.

G Protein Coupled Receptors

15 One important class of genes to which siRNAs can be directed are G-protein coupled receptors (GPCRs). GPCRs constitute a superfamily of seven transmembrane spanning proteins that respond to a diverse array of sensory and chemical stimuli, such as light, odor, taste, pheromones, hormones and neurotransmitters. GPCRs play a 20 central role in cell proliferation, differentiation, and have been implicated in the etiology of disease.

The mechanism by which G protein-coupled receptors translate extracellular signals into cellular changes was initially envisioned as a simple linear model: 25 activation of the receptor by agonist binding leads to dissociation of the heterotrimeric GTP-binding G protein (Gs, Gi, or Gq) into its alpha and beta/gamma subunits, both of which can activate or inhibit various downstream effector molecules. More specifically, activation of the GPCR induces a conformational change in the $G\alpha$ subunit, causing GDP to be released and GTP to be bound in its place. The $G\alpha$ and 30 $G\beta\gamma$ subunits then dissociate from the receptor and interact with a variety of effector molecules. For instance in the case of the Gs family, the primary function is to stimulate the intracellular messenger adenylate cyclase (AC), which catalyzes the conversion of cytoplasmic ATP into the secondary messenger cyclic AMP (cAMP). In contrast, the Gi family inhibits this pathway and the Gq family activates

phospholipases C (PLC), which cleaves phosphatidylinositol 4,5, biphosphate (PIP₂) to generate inositol-1,4,5-phosphate (IP₃) and diacylglycerol (DAG).

More recently, studies have shown that the functions of GPCRs are not limited to their actions on G-proteins and that considerable cross-talk exists between this diverse group of receptor molecules and a second class of membrane bound proteins, the receptor tyrosine kinases (RTKs). A number of GPCRs such as endothelin-1, thrombin, bombesin, and dopamine receptors can activate MAPKs, a downstream effector of the RTK/Ras pathway. Interestingly, the interaction between these two families is not unidirectional and RTKs can also modulate the activity of signaling pathways traditionally thought to be controlled exclusively by ligands that couple to GPCRs. For instance, EGF, which normally activates the MAPK cascade via the EGF receptor can stimulate adenylate cyclase activity by activating G α s.

There are dozens of members of the G Protein-Coupled Receptor family that have emerged as prominent drug targets in the last decade. One non-limiting list of potential GPCR-siRNA targets is as follows:

CMKLR1

CML1/ CMKLR1 (Accession No. Q99788) is a member of the chemokine receptor family of GPCRs that may play a role in a number of diseases including those involved in inflammation and immunological responses (*e.g.* asthma, arthritis). For this reason, siRNA directed against this protein may prove to be important therapeutic reagents.

Studies of juvenile-onset neuronal ceroid lipofuscinosis (JNCL, Batten disease), the most common form of childhood encephalopathy that is characterized by progressive neural degeneration, show that it is brought on by mutations in a novel lysosomal membrane protein (CLN3). In addition to being implicated in JNCL, CLN3 (GPCR-like protein, Accession No. A57219) expression studies have shown that the CLN3 mRNA and protein are highly over-expressed in a number of cancers (*e.g.* glioblastomas, neuroblastomas, as well as cancers of the prostate, ovaries, breast, and colon) suggesting a possible contribution of this gene to tumor growth. For this

reason, siRNA directed against this protein may prove to be important therapeutic reagents.

CLACR

5 The calcitonin receptor (CTR/ CALCR, Accession No. NM_001742) belongs to "family B" of GPCRs which typically recognized regulatory peptides such as parathyroid hormone, secretin, glucagons and vasoactive intestinal polypeptide. Although the CT receptor typically binds to calcitonin (CT), a 32 amino acid peptide hormone produced primarily by the thyroid, association of the receptor with RAMP
10 (Receptor Activity Modulating Protein) enables it to readily bind other members of the calcitonin peptide family including amylin (AMY) and other CT gene-related peptides (*e.g.* α CGRP and β CGRP). While the primary function of the calcitonin receptor pertains to regulating osteoclast mediated bone resorption and enhanced Ca^{+2} excretion by the kidney, recent studies have shown that CT and CTRs may play an
15 important role in a variety of processes as wide ranging as embryonic/foetal development and sperm function/physiology. In addition, studies have shown that patients with particular CTR genotypes may be at higher risk to lose bone mass and that this GPCR may contribute to the formation of calcium oxalate urinary stones. For this reason, siRNA directed against CTR may be useful as therapeutic reagents.

OXTR

20 The human oxytocin receptor (OTR, OXTR) is a 389 amino acid polypeptide that exhibits the seven transmembrane domain structure and belongs to the Class-I (rhodopsin-type) family of G-protein coupled receptors. OTR is expressed in a wide
25 variety of tissues throughout development and mediates physiological changes through G(q) proteins and phospholipase C-beta. Studies on the functions of oxytocin and the oxytocin receptor have revealed a broad list of duties. OT and OTR play a role in a host of sexual, maternal and social behaviors that include egg-laying, birth, milk-letdown, feeding, grooming, memory and learning. In addition, it has been
30 hypothesized that abnormalities in the functionality of oxytocin-OTR receptor-ligand system can lead to a host of irregularities including compulsive behavior, eating disorders (such as anorexia), depression, and various forms of neurodegenerative

diseases. For these reasons, siRNA directed against this gene (NM_000916) may play an important role in combating OTR-associated illnesses.

EDG GPCRs

- 5 Lysophosphatidic acid and other lipid-based hormones/growth factors induce their effects by activating signaling pathways through the G-protein coupled receptors (GPCRs) and have been observed to play important roles in a number of human diseases including cancer, asthma, and vascular pathologies. For instance, during studies of immunoglobulin A nephropathy (IgAN), researchers have observed an
- 10 enhanced expression of EDG5 (NP_004221) suggesting a contribution of this gene product in the development of IgAN. For that reasons, siRNA directed against Edg5 (NM_004230), Edg4 (NM_004720), Edg7 (Nm_012152) and related genes may play an important role in combating human disease.

15 Genes Involved in Cholesterol Signaling and Biosynthesis

- Studies on model genetic organisms such as *Drosophila* and *C. elegans* have led to the identification of a plethora of genes that are essential for early development. Mutational analysis and ectopic expression studies have allowed many of these genes to be grouped into discreet signal transduction pathways and have shown that these
- 20 elements play critical roles in pattern formation and cell differentiation. Disruption of one or more of these genes during early stages of development frequently leads to birth defects whereas as alteration of gene function at later stages in life can result in tumorigenesis.

- 25 One critical set of interactions known to exist in both invertebrates and vertebrates is the Sonic Hedgehog-Patched-Gli pathway. Originally documented as a *Drosophila* segmentation mutant, several labs have recently identified human and mouse orthologs of many of the pathways members and have successfully related disruptions in these genes to known diseases. Pathway activation is initiated with the
- 30 secretion of Sonic hedgehog. There are three closely related members of the Shh family (Sonic hedgehog, Desert, and Indian) with Shh being the most widely expressed form of the group. The Shh gene product is secreted as a small pro-signal molecule. To successfully initiate its developmental role, Shh is first cleaved, whereupon the N-terminal truncated fragment is covalently modified with cholesterol.

The addition of the sterol moiety promotes the interaction between Shh and its cognate membrane bound receptor, Patched (Ptch). There are at least two isoforms of the Patched gene, Ptch1 and Ptch2. Both isoforms contain a sterol-sensing domain (SSD); a roughly 180 amino acid cluster that is found in at least seven different classes of molecules including those involved in cholesterol biosynthesis, vesicular traffic, signal transduction, cholesterol transport, and sterol homeostasis. In the absence of Shh, the Patched protein is a negative regulator of the pathway. In contrast, binding of Shh-cholesterol to the Patched receptor releases the negative inhibition which that molecule enforces on a G-protein coupled receptor known as Smoothened. Subsequent activation of Smoothened (directly or indirectly) leads to the triggering of a trio of transcription factors that belong to the Gli family. All three factors are relatively large, contain a characteristic C2-H2 zinc-finger-pentamer, and recognize one of two consensus sequences (SEQ. ID NO. 0463 GACCACCCA or SEQ. ID NO. 0464 GAACCACCCA). In the absence of Shh, Gli proteins are cleaved by the proteasome and the C-terminally truncated fragment translocates to the nucleus and acts as a dominant transcription repressor. In the presence of Shh-cholesterol, Gli repressor formation is inhibited and full-length Gli functions as a transcriptional activator.

Shh and other members of the Shh-PTCH-Gli pathway are expressed in a broad range of tissues (*e.g.* the notochord, the floorplate of the neural tube, the brain, and the gut) at early stages in development. Not surprisingly, mutations that lead to altered protein expression or function have been shown to induce developmental abnormalities. Defects in the human Shh gene have been shown to cause holoprosencephaly, a midline defect that manifests itself as cleft lip or palate, CNS septation, and a wide range of other phenotypes. Interestingly, defects in cholesterol biosynthesis generate similar Shh-like disorders (*e.g.* Smith-Lemli-Opitz syndrome) suggesting that cholesterol modification of the Shh gene product is crucial for pathway function. Both the Patched and Smoothened genes have also been shown to be clinically relevant with Smoothened now being recognized as an oncogene that, like PTCH-1 and PTCH-2, is believed to be the causative agent of several forms of adult tumors. For these reasons, siRNA directed against Smoothened (SMO, NM_005631), Patched (PTCH, nm_000264), and additional genes that participate in

cholesterol signaling, biosynthesis, and degradation, have potentially useful research and therapeutic applications.

Targeted Pathways.

- 5 In addition to targeting siRNA against one or more members of a family of proteins, siRNA can be directed against members of a pathway. Thus, for instance, siRNA can be directed against members of a signal transduction pathway (*e.g.* the insulin pathway, including AKT1-3, CBL, CBLB, EIF4EBP1, FOXO1A, FOXO3A, FRAP1, GSK3A, GSK3B, IGF1, IGF1R, INPP5D, INSR, IRS1, MLLT7, PDPK1,
- 10 PIK3CA, PIK3CB, PIK3R1, PIK3R2, PPP2R2B, PTEN, RPS6, RPS6KA1, RPX6KA3, SGK, TSC1, TSC2, AND XPO1), an apoptotic pathway (CASP3,6,7,8,9, DSH1/2, P110, P85, PDK1/2, CATENIN, HSP90, CDC37, P23, BAD, BCLXL, BCL2, SMAC, and others), pathways, involved in DNA damage, cell cycle, and other physiological (p53,MDM2, CHK1/2, BRCA1/2, ATM, ATR, P15INK4, P27, P21,
- 15 SKP2, CDC25C/A, 14-3-3, PLK, RB, CDK4, GLUT4, Inos, Mtor, FKBP, PPAR, RXR, ER). Similarly, genes involved in immune system function including TNFR1, IL-IR, IRAK1/2, TRAF2, TRAF6, TRADD, FADD, IKK ϵ , IKK γ , IKK β , IKK α , Ikb α , Ikb β , p50, p65, Rac, RhoA, Cdc42, ROCK, Pak1/2/3/4/5/6, cIAP, HDAC1/2, CBP, β -TrCP, Rip2/4, and others are also important targets for the siRNAs described
- 20 in this document and may be useful in treating immune system disorders. Genes involved in apoptosis, such as Dsh1/2,PTEN, P110 (pan), P85, PDK1/2, Akt1, Akt2, Akt (pan), p70^{S6K}, GSK3 β , PP2A (cat), β -catenin, HSP90, Cdc37/p50, P23, Bad, BclxL, Bcl2, Smac/Diablo, and Ask1 are potentially useful in the treatment of diseases that involve defects in programmed cell death (*e.g.* cancer), while siRNA
- 25 agents directed against p53, MDM2, Chk1/2, BRCA1/2, ATM, ATR, p15^{INK4}, P27, P21, Skp2, Cdc25C/A, 14-3-3 σ/ϵ , PLK, Rb, Cdk4, Glut4, iNOS, mTOR, FKBP, PPAR γ , RXR α , ER α and related genes may play a critical role in combating diseases associated with disruptions in DNA repair, and cell cycle abnormalities.
- 30 Tables VI -Table X below provide examples of useful pools for inhibiting different genes in the human insulin pathway and tyrosine kinase pathways, proteins involved in the cell cycle, the production of nuclear receptors, and other genes. These particular pools are particularly useful in humans, but would be useful in any species

that generates an appropriately homologous mRNA. Further, within each of the listed pools any one sequence maybe used independently but preferably at least two of the listed sequences, more preferably at least three, and most preferably all of the listed sequences for a given gene is present.

5

Table VI

Gene Name	Acc#	GI	L.L.	Duplex #	Sequence	SEQ. ID NO
AKT1	NM 005163	4885060	207	D-003000-05	GACAAGGACGGGCACATTA	465
AKT1	NM 005163	4885060	207	D-003000-06	GGACAAGGACGGGCACATT	466
AKT1	NM 005163	4885060	207	D-003000-07	GCTACTTCCTCCTCAAGAA	467
AKT1	NM 005163	4885060	207	D-003000-08	GACCGCCTCTGCTTTGTCA	468
AKT2						
AKT2	NM 001626	6715585	208	D-003001-05	GTACTTCGATGATGAATTT	469
AKT2	NM 001626	6715585	208	D-003001-06	GCAAAGAGGGCATCAGTGA	470
AKT2	NM 001626	6715585	208	D-003001-07	GGGCTAAAGTGACCATGAA	471
AKT2	NM 001626	6715585	208	D-003001-08	GCAGAATGCCAGCTGATGA	472
AKT3						
AKT3	NM 005465	32307164	10000	D-003002-05	GGAGTAAACTGGCAAGATG	473
AKT3	NM 005465	32307164	10000	D-003002-06	GACATTAAATTTCTCGAA	474
AKT3	NM 005465	32307164	10000	D-003002-07	GACCAAAGCCAAACACATT	475
AKT3	NM 005465	32307164	10000	D-003002-08	GAGGAGAGAATGAATTGTA	476
CBL						
CBL	NM 005188	4885116	867	D-003003-05	GGAGACACATTTCTGGATTA	477
CBL	NM 005188	4885116	867	D-003003-06	GATCTGACCTGCAATGATT	478
CBL	NM 005188	4885116	867	D-003003-07	GACAATCCCTCACAATAAA	479
CBL	NM 005188	4885116	867	D-003003-08	CCAGAAAGCTTTGTCATT	480
CBLB						
CBLB	NM 170662	29366807	868	D-003004-05	GACCATACCTCATAACAAG	481
CBLB	NM 170662	29366807	868	D-003004-06	TGAAAGACCTCCACCAATC	482
CBLB	NM 170662	29366807	868	D-003004-07	GATGAAGGCTCCAGGTGTT	483
CBLB	NM 170662	29366807	868	D-003004-08	TATCAGCATTACGACTTA	484
EIF4EBP1						
EIF4EBP1	NM 004095	20070179	1978	D-003005-05	GCAATAGCCCAGAAGATAA	485
EIF4EBP1	NM 004095	20070179	1978	D-003005-06	CGCAATAGCCCAGAAGATA	486
EIF4EBP1	NM 004095	20070179	1978	D-003005-07	GAGATGGACATTTAAAGCA	487
EIF4EBP1	NM 004095	20070179	1978	D-003005-08	CAATAGCCCAGAAGATAAG	488
FOXO1A						
FOXO1A	NM 002015	9257221	2308	D-003006-05	CCAGGCATCTCATAACAAA	489
FOXO1A	NM 002015	9257221	2308	D-003006-06	CCAGATGCCTATACAAACA	490
FOXO1A	NM 002015	9257221	2308	D-003006-07	GGAGGTATGAGTCAGTATA	491
FOXO1A	NM 002015	9257221	2308	D-003006-08	GAGGTATGAGTCAGTATAA	492
FOXO3A						
FOXO3A	NM 001455	4503738	2309	D-003007-01	CAATAGCAACAAGTATACC	493
FOXO3A	NM 001455	4503738	2309	D-003007-02	TGAAGTCCAGGACGATGAT	494
FOXO3A	NM 001455	4503738	2309	D-003007-03	TGTCACACTATGGTAACCA	495
FOXO3A	NM 001455	4503738	2309	D-003007-04	TGTTCAATGGGAGCTTGA	496
FRAP1						
FRAP1	NM 004958	19924298	2475	D-003008-05	GAGAAGAAATGGAAGAAAT	497
FRAP1	NM 004958	19924298	2475	D-003008-06	CCAAAGTGCTGCAGTACTA	498
FRAP1	NM 004958	19924298	2475	D-003008-07	GAGCATGCCGTCAATAATA	499
FRAP1	NM 004958	19924298	2475	D-003008-08	GGTCTGAAGTGAATGAAGA	500

GSK3A						
GSK3A	NM_019884	11995473	2931	D-003009-05	GGACAAAGGTGTTCAAATC	501
GSK3A	NM_019884	11995473	2931	D-003009-06	GAACCCAGCTGCCTAACAA	502
GSK3A	NM_019884	11995473	2931	D-003009-07	GCGCACAGCTTCTTTGATG	503
GSK3A	NM_019884	11995473	2931	D-003009-08	GCTCTAGCCTGCTGGAGTA	504
GSK3B						
GSK3B	NM_002093	21361339	2932	D-003010-05	GAAGAAAGATGAGGTCTAT	505
GSK3B	NM_002093	21361339	2932	D-003010-06	GGACCCAAATGTCAAACATA	506
GSK3B	NM_002093	21361339	2932	D-003010-07	GAAATGAACCCAACTACA	507
GSK3B	NM_002093	21361339	2932	D-003010-08	GATGAGGTCTATCTTAATC	508
IGF1						
IGF1	NM_000618			D-003011-05	GGAAGTACATTTGAAGAAC	509
IGF1	NM_000618			D-003011-06	AGAAGGAAGTACATTTGAA	510
IGF1	NM_000618			D-003011-07	CCTCAAGCCTGCCAAGTCA	511
IGF1	NM_000618			D-003011-08	GGTGGATGCTCTTCAGTTC	512
IGF1R						
IGF1R	NM_000875	11068002	3480	D-003012-05	CAACGAAGCTTCTGTGATG	513
IGF1R	NM_000875	11068002	3480	D-003012-06	GGCCAGAAATGGAGAATAA	514
IGF1R	NM_000875	11068002	3480	D-003012-07	GAAGCACCCCTTTAAGAATG	515
IGF1R	NM_000875	11068002	3480	D-003012-08	GCAGACACCTACAACATCA	516
INPP5D						
INPP5D	NM_005541	5031798	3635	D-003013-05	GGAATTGCGTTTACACTTA	517
INPP5D	NM_005541	5031798	3635	D-003013-06	GGAAACTGATCATTAAGAA	518
INPP5D	NM_005541	5031798	3635	D-003013-07	CGACAGGGATGAAGTACAA	519
INPP5D	NM_005541	5031798	3635	D-003013-08	AAACGCAGCTGCCCATCTA	520
INSR						
INSR	NM_000208	4557883	3643	D-003014-05	GGAAGACGTTTGAGGATTA	521
INSR	NM_000208	4557883	3643	D-003014-06	GAACAAGGCTCCCGAGAGT	522
INSR	NM_000208	4557883	3643	D-003014-07	GGAGAGACCTTGGAAATTG	523
INSR	NM_000208	4557883	3643	D-003014-08	GGACGGAACCCACCTATTT	524
IRS1						
IRS1	NM_005544	5031804	3667	D-003015-05	AAAGAGGTCTGGCAAGTGA	525
IRS1	NM_005544	5031804	3667	D-003015-06	GAACCTGATTGGTATCTAC	526
IRS1	NM_005544	5031804	3667	D-003015-07	CCACGGCGATCTAGTGCTT	527
IRS1	NM_005544	5031804	3667	D-003015-08	GTCAGTCTGTCGTCCAGTA	528
MLLT7						
MLLT7	NM_005938	5174578	4303	D-003016-05	GGAAGGACTTCAACTTTG	529
MLLT7	NM_005938	5174578	4303	D-003016-06	CCACGAAGCAGTTCAAATG	530
MLLT7	NM_005938	5174578	4303	D-003016-07	GAGAAGCGACTGACACTTG	531
MLLT7	NM_005938	5174578	4303	D-003016-08	GACCAGAGATCGCTAACCA	532
PDPK1						
PDPK1	NM_002613	4505694	5170	D-003017-05	CAAGAGACCTCGTGGAGAA	533
PDPK1	NM_002613	4505694	5170	D-003017-06	GACCAGAGGCCAAGAATT	534
PDPK1	NM_002613	4505694	5170	D-003017-07	GGAAACGAGTATCTTATAT	535
PDPK1	NM_002613	4505694	5170	D-003017-08	GAGAAGCGACATATCATAA	536
PIK3CA						
PIK3CA	NM_006218	5453891	5290	D-003018-05	GCTATCATCTGAACAATTA	537
PIK3CA	NM_006218	5453891	5290	D-003018-06	GGATAGAGGCCAATAATA	538
PIK3CA	NM_006218	5453891	5290	D-003018-07	GGACAACTGTTTCATATAG	539
PIK3CA	NM_006218	5453891	5290	D-003018-08	GCCAGTACCTCATGGATTA	540
PIK3CB						
PIK3CB	NM_006219	5453893	5291	D-003019-05	CGACAAGACTGCCGAGAGA	541
PIK3CB	NM_006219	5453893	5291	D-003019-06	TCAAGTGTCTCCTAATATG	542
PIK3CB	NM_006219	5453893	5291	D-003019-07	GGATTGAGTTGGAGTGATT	543
PIK3CB	NM_006219	5453893	5291	D-003019-08	TTTCAAGTGTCTCCTAATA	544
PIK3R1						

PIK3R1	NM_181504	32455251	5295	D-003020-05	GGAAATATGGCTTCTCTGA	545
PIK3R1	NM_181504	32455251	5295	D-003020-06	GAAAGACGAGAGACCAATA	546
PIK3R1	NM_181504	32455251	5295	D-003020-07	GTAAAGCATTGTGTCATAA	547
PIK3R1	NM_181504	32455251	5295	D-003020-08	GGATCAAGTTGTCAAAGAA	548
PIK3R2						
PIK3R2	NM_005027	4826907	5296	D-003021-05	GGAAAGGCGGGAACAATAA	549
PIK3R2	NM_005027	4826907	5296	D-003021-06	GATGAAGCGTACTGCAATT	550
PIK3R2	NM_005027	4826907	5296	D-003021-07	GGACAGCGAATCTCACTAC	551
PIK3R2	NM_005027	4826907	5296	D-003021-08	GCAAGATCCGAGACCAGTA	552
PPP2R2B						
PPP2R2B	NM_004576	4758953	5521	D-003022-05	GAATGCAGCTTACTTTCTT	553
PPP2R2B	NM_004576	4758953	5521	D-003022-06	GACCGAAGCTGACATTATC	554
PPP2R2B	NM_004576	4758953	5521	D-003022-07	TCGATTACCTGAAGAGTTT	555
PPP2R2B	NM_004576	4758953	5521	D-003022-08	CCTGAAGAGTTTAGAAATA	556
PTEN						
PTEN	NM_000314	4506248	5728	D-003023-05	GTGAAGATCTTGACCAATG	557
PTEN	NM_000314	4506248	5728	D-003023-06	GATCAGCATACACAAATTA	558
PTEN	NM_000314	4506248	5728	D-003023-07	GGCGCTATGTGTATTATTA	559
PTEN	NM_000314	4506248	5728	D-003023-08	GTATAGAGCGTGCAGATAA	560
RPS6						
RPS6	NM_001010	17158043	6194	D-003024-05	GCCAGAAACTCATTGAAGT	561
RPS6	NM_001010	17158043	6194	D-003024-06	GGATATTCCTGGACTGACT	562
RPS6	NM_001010	17158043	6194	D-003024-07	CCAAGGAGAACTGGAGAAA	563
RPS6	NM_001010	17158043	6194	D-003024-08	GCGTATGGCCACAGAAGTT	564
RPS6KA1						
RPS6KA1	NM_002953	20149546	6195	D-003025-05	GATGACACCTTCTACTTTG	565
RPS6KA1	NM_002953	20149546	6195	D-003025-06	GAGAATGGGCTCCTCATGA	566
RPS6KA1	NM_002953	20149546	6195	D-003025-07	CAAGCGGGATCCTTCAGAA	567
RPS6KA1	NM_002953	20149546	6195	D-003025-08	CCACCGGCCTGATGGAAGA	568
RPS6KA3						
RPS6KA3	NM_004586	4759049	6197	D-003026-05	GAAGGGAAGTTGTATCTTA	569
RPS6KA3	NM_004586	4759049	6197	D-003026-06	GAAAGTATGTGTATGTAGT	570
RPS6KA3	NM_004586	4759049	6197	D-003026-07	GGACAGCATCCAAACATTA	571
RPS6KA3	NM_004586	4759049	6197	D-003026-08	GGAGGTGAATTGCTGGATA	572
SGK						
SGK	NM_005627	5032090	6446	D-003027-01	TTAATGGTGGAGAGTTGTT	573
SGK	NM_005627	5032090	6446	D-003027-04	ATTAAGTGGGATGATCTCA	574
SGK	NM_005627	25168262	6446	D-003027-05	GAAGAAAGCAATCCTGAAA	575
SGK	NM_005627	25168262	6446	D-003027-06	AAACACAGCTGAAATGTAC	576
TSC1						
TSC1	NM_000368	24475626	7248	D-003028-05	GAAGATGGCTATTCTGTGT	577
TSC1	NM_000368	24475626	7248	D-003028-06	TATGAAGGCTCGAGAGTTA	578
TSC1	NM_000368	24475626	7248	D-003028-07	CGACACGGCTGATAACTGA	579
TSC1	NM_000368	24475626	7248	D-003028-08	CGGCTGATGTTGTTAAATA	580
TSC2						
TSC2	NM_000548	10938006	7249	D-003029-05	GCATTAATCTCTTACCATA	581
TSC2	NM_000548	10938006	7249	D-003029-06	CCAATGTCCTCTTGTCTTT	582
TSC2	NM_000548	10938006	7249	D-003029-07	GGAGACACATCACCTACTT	583
TSC2	NM_000548	10938006	7249	D-003029-08	TCACCAGGCTCATCAAGAA	584
XPO1						
XPO1	NM_003400	8051634	7514	D-003030-05	GAAAGTCTCTGTCAAATA	585
XPO1	NM_003400	8051634	7514	D-003030-06	GCAATAGGCTCCATTAGTG	586
XPO1	NM_003400	8051634	7514	D-003030-07	GGAACATGATCAACTTATA	587
XPO1	NM_003400	8051634	7514	D-003030-08	GGATACAGATTCCATAAAT	588

Table VII

Gene Name	Acc#	GI	L.L	Duplex #	Sequence	St ID
ABL1						
ABL1	NM_007313	6382057	25	D-003100-05	GGAAATCAGTGACATAGTG	5
ABL1	NM_007313	6382057	25	D-003100-06	GGTCCACACTGCAATGTTT	5
ABL1	NM_007313	6382057	25	D-003100-07	GAAGGAAATCAGTGACATA	5
ABL1	NM_007313	6382057	25	D-003100-08	TCACTGAGTTCATGACCTA	5
ABL2						
ABL2	NM_007314	6382061	27	D-003101-05	GAAATGGAGCGAACAGATA	5
ABL2	NM_007314	6382061	27	D-003101-06	GAGCCAAATTTCTATTAA	5
ABL2	NM_007314	6382061	27	D-003101-07	GTAATAAGCCTACAGTCTA	5
ABL2	NM_007314	6382061	27	D-003101-08	GGAGTGAAGTTCGCTCTAA	5
ACK1						
ACK1	NM_005781	8922074	10188	D-003102-05	AAACGCAAGTCGTGGATGA	5
ACK1	NM_005781	8922074	10188	D-003102-06	GCAAGTCGTGGATGAGTAA	5
ACK1	NM_005781	8922074	10188	D-003102-07	GAGCACTACCTCAGAATGA	5
ACK1	NM_005781	8922074	10188	D-003102-08	TCAGCAGCACCCACTATTA	6
ALK						
ALK	NM_004304	29029631	238	D-003103-05	GACAAGATCCTGCAGAATA	6
ALK	NM_004304	29029631	238	D-003103-06	GGAAGAGTCTGGCAGTTGA	6
ALK	NM_004304	29029631	238	D-003103-07	GCACGTGGCTCGGGACATT	6
ALK	NM_004304	29029631	238	D-003103-08	GAACTGCAGTGAAGGAACA	6
AXL						
AXL	NM_021913	21536465	558	D-003104-05	GGTCAGAGCTGGAGGATTT	6
AXL	NM_021913	21536465	558	D-003104-06	GAAAGAAGGAGACCCGTTA	6
AXL	NM_021913	21536465	558	D-003104-07	CCAAGAAGATCTACAATGG	6
AXL	NM_021913	21536465	558	D-003104-08	GAACTGCATGCTGAATGA	6
BLK						
BLK	NM_001715	4502412	640	D-003105-05	GAGGATGCCTGCTGGATTT	6
BLK	NM_001715	4502412	640	D-003105-06	ACATGAAGGTGGCCATTAA	6
BLK	NM_001715	4502412	640	D-003105-07	GGTCAGCGCCCAAGACAAG	6
BLK	NM_001715	4502412	640	D-003105-08	GAACTCGGGTCTGGACAA	6
BMX						
BMX	NM_001721	21359831	660	D-003106-05	AAACAAACCTTTCTACTA	6
BMX	NM_001721	21359831	660	D-003106-06	GAAGGAGCATTTATGGTTA	6
BMX	NM_001721	21359831	660	D-003106-07	GAGAAGAGATTACCTTGTT	6
BMX	NM_001721	21359831	660	D-003106-08	GTAAGGCTGTGAATGATAA	6
BTK						
BTK	NM_000061	4557376	695	D-003107-05	GAACAGGAATGGAAGCTTA	6
BTK	NM_000061	4557376	695	D-003107-06	GCTATGGGCTGCCAAATTT	6
BTK	NM_000061	4557376	695	D-003107-07	GAAAGCAACTTACCATGGT	6
BTK	NM_000061	4557376	695	D-003107-08	GGTAAACGATCAAGGAGTT	6
C20orf64						
C20orf64	NM_033550	19923655	11285	D-003108-05	CAACTTAGCCAAGACAATT	6
C20orf64	NM_033550	19923655	11285	D-003108-06	GAAATTGAAGGCTCAGTGA	6
C20orf64	NM_033550	19923655	11285	D-003108-07	TGGAACAGCTGAACATTGT	6
C20orf64	NM_033550	19923655	11285	D-003108-08	GCTTCCAAGTCTTATATA	6
CSF1R						
CSF1R	NM_005211	27262658	1436	D-003109-05	GGAGAGCTCTGACGTTTGA	6
CSF1R	NM_005211	27262658	1436	D-003109-06	CAACAACGCTACCTTCCAA	6
CSF1R	NM_005211	27262658	1436	D-003109-07	CCACGCAGCTGCCTTACAA	6
CSF1R	NM_005211	27262658	1436	D-003109-08	GGAACAACCTGCAGTTTGG	6
CSK						

CSK	NM_004383	4758077	1445	D-003110-05	CAGAATGTATTGCCAAGTA	6
CSK	NM_004383	4758077	1445	D-003110-06	GAACAAAGTCGCCGTCAAG	6
CSK	NM_004383	4758077	1445	D-003110-07	GCGAGTGCCTTATCCAAGA	6
CSK	NM_004383	4758077	1445	D-003110-08	GGAGAAGGGCTACAAGATG	6
DDR1						
DDR1	NM_013994	7669484	780	D-003111-05	GGAGATGGAGTTTGAGTTT	6
DDR1	NM_013994	7669484	780	D-003111-06	CAGAGGCCCTGTCATCTTT	6
DDR1	NM_013994	7669484	780	D-003111-07	GCTGGTAGCTGTCAAGATC	6
DDR1	NM_013994	7669484	780	D-003111-08	TGAAAGAGGTGAAGATCAT	6
DDR2						
DDR2	NM_006182	5453813	4921	D-003112-05	GGTAAGAACTACACAATCA	6
DDR2	NM_006182	5453813	4921	D-003112-06	GAACGAGAGTGCCACCAAT	6
DDR2	NM_006182	5453813	4921	D-003112-07	ACACCAATCTGAAGTTTAT	6
DDR2	NM_006182	5453813	4921	D-003112-08	CAACAAGAATGCCAGGAAT	6
DKFZp761 P1010						
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-05	CCTAGAAGCTGCCATTAAA	6
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-06	GATTAGGCCTGGCTTATGA	6
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-07	CCCAGTAGCTGCACACATA	6
DKFZp761 P1010	NM_018423	8922178	55359	D-003113-08	GGTGGTACCTGAACTGTAT	6
EGFR						
EGFR	NM_005228	4885198	1956	D-003114-05	GAAGGAACTGAATTCAAA	6
EGFR	NM_005228	4885198	1956	D-003114-06	GGAAATATGTACTACGAAA	6
EGFR	NM_005228	4885198	1956	D-003114-07	CCACAAAGCAGTGAATTTA	6
EGFR	NM_005228	4885198	1956	D-003114-08	GTAACAAGCTCACGCAGTT	6
EPHA1						
EPHA1	NM_005232	4885208	2041	D-003115-05	GACCAGAGCTTCACCATTC	6
EPHA1	NM_005232	4885208	2041	D-003115-06	GCAAGACTGTGGCCATTAA	6
EPHA1	NM_005232	4885208	2041	D-003115-07	GGGCGAACCTGACCTATGA	6
EPHA1	NM_005232	4885208	2041	D-003115-08	GATTGTAGCCGTCTCTTT	6
EPHA2						
EPHA2	NM_004431	4758277	1969	D-003116-05	GGAGGGATCTGGCAACTTG	6
EPHA2	NM_004431	4758277	1969	D-003116-06	GCAGCAAGGTGCACGAATT	6
EPHA2	NM_004431	4758277	1969	D-003116-07	GGAGAAGGATGGCGAGTTC	6
EPHA2	NM_004431	4758277	1969	D-003116-08	GAAGTTCACTACCGAGATC	6
EPHA3						
EPHA3	NM_005233	21361240	2042	D-003117-05	GATCGGACCTCCAGAAATA	6
EPHA3	NM_005233	21361240	2042	D-003117-06	GAAGCTCAGCTCAGAAGATT	6
EPHA3	NM_005233	21361240	2042	D-003117-07	GCAAGAGGCACAAATGTTA	6
EPHA3	NM_005233	21361240	2042	D-003117-08	GAGCATCAGTTTACAAAGA	6
EPHA4						
EPHA4	NM_004438	4758279	2043	D-003118-05	GGTCTGGGATGAAGTATTT	6
EPHA4	NM_004438	4758279	2043	D-003118-06	GAATGAAGTTACCTTATTG	6
EPHA4	NM_004438	4758279	2043	D-003118-07	GAAGTTGGGTGGATAGCAA	6
EPHA4	NM_004438	4758279	2043	D-003118-08	GAGATTAAATTCACCTTGA	6
EPHA7						
EPHA7	NM_004440	4758281	2045	D-003119-05	GAAAAGAGATGTTGCAGTA	6
EPHA7	NM_004440	4758281	2045	D-003119-06	CTAGATGCCTCCTGTATTA	6
EPHA7	NM_004440	4758281	2045	D-003119-07	AGAAGAAGGTTATCGTTTA	6
EPHA7	NM_004440	4758281	2045	D-003119-08	TAGCAAAGCTGACCAAGAA	6
EPHA8						
EPHA8	NM_020526	18201903	2046	D-003120-05	GAAGATGCACTATCAGAAT	6

EPHA8	NM 020526	18201903	2046	D-003120-06	GAGAAGATGCACTATCAGA	6
EPHA8	NM 020526	18201903	2046	D-003120-07	AACCTGATCTCCAGTGTGA	6
EPHA8	NM 020526	18201903	2046	D-003120-08	TCTCAGACCTGGGCTATGT	6
EPHB1						
EPHB1	NM 004441	21396502	2047	D-003121-05	GCGATAAGCTCCAGCATT	6
EPHB1	NM 004441	21396502	2047	D-003121-06	GAAACGGGCTTATAGCAA	6
EPHB1	NM 004441	21396502	2047	D-003121-07	GGATGAAGATCTACATTGA	6
EPHB1	NM 004441	21396502	2047	D-003121-08	GCACGTCTCTGTCAACATC	6
EPHB2						
EPHB2	NM 017449	17975764	2048	D-003122-05	ACTATGAGCTGCAGTACTA	6
EPHB2	NM 017449	17975764	2048	D-003122-06	GTACAACGCCACAGCCATA	6
EPHB2	NM 017449	17975764	2048	D-003122-07	GGAAAGCAATGACTGTTCT	6
EPHB2	NM 017449	17975764	2048	D-003122-08	CGGACAAGCTGCAACACTA	6
EPHB3						
EPHB3	NM 004443	17975767	2049	D-003123-05	GGTGTGATCTCCAATGTGA	6
EPHB3	NM 004443	17975767	2049	D-003123-06	GGGATGACCTCCTGTACAA	6
EPHB3	NM 004443	17975767	2049	D-003123-07	CAGAAGACCTGCTCCGTAT	6
EPHB3	NM 004443	17975767	2049	D-003123-08	GAGATGAAGTACTTTGAGA	6
EPHB4						
EPHB4	NM 004444	17975769	2050	D-003124-05	GGACAAACACGGACAGTAT	6
EPHB4	NM 004444	17975769	2050	D-003124-06	GTACTAAGGTCTACATCGA	6
EPHB4	NM 004444	17975769	2050	D-003124-07	GGAGAGAAGCAGAATATTC	6
EPHB4	NM 004444	17975769	2050	D-003124-08	GCCAATAGCCACTCTAACA	6
EPHB6						
EPHB6	NM 004445	4758291	2051	D-003125-05	GGAAGTCGATCCTGCTTAT	6
EPHB6	NM 004445	4758291	2051	D-003125-06	GGACCAAGGTGGACACAAT	6
EPHB6	NM 004445	4758291	2051	D-003125-07	TGTGGGAAGTGATGAGTTA	6
EPHB6	NM 004445	4758291	2051	D-003125-08	CGGGAGACCTTCACCCTTT	6
ERBB2						
ERBB2	NM 004448	4758297	2064	D-003126-05	GGACGAATTCTGCACAATG	6
ERBB2	NM 004448	4758297	2064	D-003126-06	GACGAATTCTGCACAATGG	6
ERBB2	NM 004448	4758297	2064	D-003126-07	CTACAACACAGACACGTTT	6
ERBB2	NM 004448	4758297	2064	D-003126-08	AGACGAAGCATACGTGATG	6
ERBB3						
ERBB3	NM 001982	4503596	2065	D-003127-05	AAGAGGATGTCAACGGTTA	6
ERBB3	NM 001982	4503596	2065	D-003127-06	GAAGACTGCCAGACATTGA	6
ERBB3	NM 001982	4503596	2065	D-003127-07	GACAAACACTGGTGCTGAT	6
ERBB3	NM 001982	4503596	2065	D-003127-08	GCAGTGGATTTCGAGAAGTG	6
ERBB4						
ERBB4	NM 005235	4885214	2066	D-003128-05	GAGGAAAGATGCCAATTAA	6
ERBB4	NM 005235	4885214	2066	D-003128-06	GCAGGAAACATCTATATTA	6
ERBB4	NM 005235	4885214	2066	D-003128-07	GATCACAACCTGCTGCTTAA	6
ERBB4	NM 005235	4885214	2066	D-003128-08	CCTCAAAGATACCTAGTTA	6
FER						
FER	NM 005246	4885230	2241	D-003129-05	GGAGTGACCTGAAGAATTC	6
FER	NM 005246	4885230	2241	D-003129-06	TAAAGCAGATTCCCATTAA	6
FER	NM 005246	4885230	2241	D-003129-07	GGAAAGTACTGTCCAAATG	6
FER	NM 005246	4885230	2241	D-003129-08	GAACAACGGCTGCTAAAGA	6
FES						
FES	NM 002005	13376997	2242	D-003130-05	CGAGGATCCTGAAGCAGTA	6
FES	NM 002005	13376997	2242	D-003130-06	AGGAATACCTGGAGATTAG	6
FES	NM 002005	13376997	2242	D-003130-07	CAACAGGAGCTCCGGAATG	6
FES	NM 002005	13376997	2242	D-003130-08	GGTGTGGGTGAGCAGATT	6
FGFR1						
FGFR1	NM 000604	13186232	2260	D-003131-05	TAAGAAATGTCTCCTTTGA	6
FGFR1	NM 000604	13186232	2260	D-003131-06	GAAGACTGCTGGAGTTAAT	6

FGFR1	NM_000604	13186232	2260	D-003131-07	GATGGTCCCTTGATGTCA	7
FGFR1	NM_000604	13186232	2260	D-003131-08	CTTAAGAAATGTCTCCTTT	7
FGFR2						
FGFR2	NM_000141	13186239	2263	D-003132-05	CCAAATCTCTCAACCAGAA	7
FGFR2	NM_000141	13186239	2263	D-003132-06	GAACAGTATTCACCTAGTT	7
FGFR2	NM_000141	13186239	2263	D-003132-07	GGCCAACACTGTCAAGTTT	7
FGFR2	NM_000141	13186239	2263	D-003132-08	GTGAAGATGTTGAAAGATG	7
FGFR3						
FGFR3	NM_000142	13112046	2261	D-003133-05	TGTCGGACCTGGTGTCTGA	7
FGFR3	NM_000142	13112046	2261	D-003133-06	GCATCAAGCTGCGGCATCA	7
FGFR3	NM_000142	13112046	2261	D-003133-07	GGACGGCACACCCTACGTT	7
FGFR3	NM_000142	13112046	2261	D-003133-08	TGCACAACCTCGACTACTA	7
FGFR4						
FGFR4	NM_002011	13112051	2264	D-003134-05	GCACTGGAGTCTCGTGATG	7
FGFR4	NM_002011	13112051	2264	D-003134-06	CATAGGGACCTCTCGAATA	7
FGFR4	NM_002011	13112051	2264	D-003134-07	ATACGGACATCATCCTGTA	7
FGFR4	NM_002011	13112051	2264	D-003134-08	ATAGGGACCTCTCGAATAG	7
FGR						
FGR	NM_005248	4885234	2268	D-003135-05	GCGATCATGTGAAGCATT	7
FGR	NM_005248	4885234	2268	D-003135-06	TCACTGAGCTCATACCAA	7
FGR	NM_005248	4885234	2268	D-003135-07	GAAGAGTGGTACTTTGGAA	7
FGR	NM_005248	4885234	2268	D-003135-08	CCCAGAAGCTGCCCTCTTT	7
FLT1						
FLT1	NM_002019	4503748	2321	D-003136-05	GAGCAAACGTGACTTATTT	7
FLT1	NM_002019	4503748	2321	D-003136-06	CCAAATGGGTTTCATGTTA	7
FLT1	NM_002019	4503748	2321	D-003136-07	CAACAAGGATGCAGCACTA	7
FLT1	NM_002019	4503748	2321	D-003136-08	GGACGTAACCTGAAGAGGAT	7
FLT3						
FLT3	NM_004119	4758395	2322	D-003137-05	GAAGGCATCTACACCATT	7
FLT3	NM_004119	4758395	2322	D-003137-06	GAAGGAGTCTGGAATAGAA	7
FLT3	NM_004119	4758395	2322	D-003137-07	GAATTTAAGTCGTGTGTTT	7
FLT3	NM_004119	4758395	2322	D-003137-08	GGAATTCATTTCACTCTGA	7
FLT4						
FLT4	NM_002020	4503752	2324	D-003138-05	GCAAGAACGTGCATCTGTT	7
FLT4	NM_002020	4503752	2324	D-003138-06	GCGAATACCTGTCTTACGA	7
FLT4	NM_002020	4503752	2324	D-003138-07	GAAGACATTTGAGGAATTC	7
FLT4	NM_002020	4503752	2324	D-003138-08	GAGCAGCCATTATCAACA	7
FRK						
FRK	NM_002031	4503786	2444	D-003139-05	GAAACAGACTCTTCATATT	7
FRK	NM_002031	4503786	2444	D-003139-06	GAACAATACCACTCCAGTA	7
FRK	NM_002031	4503786	2444	D-003139-07	CAAGACCGTTCTTTCTA	7
FRK	NM_002031	4503786	2444	D-003139-08	GCAAGAATATCTCCAAAT	7
FYN						
FYN	NM_002037	23510344	2534	D-003140-05	GGAATGGACTCATATGCAA	7
FYN	NM_002037	23510344	2534	D-003140-06	GCAGAAGAGTGGTACTTTG	7
FYN	NM_002037	23510344	2534	D-003140-07	CAAAGGAAGTTTACTGGAT	7
FYN	NM_002037	23510344	2534	D-003140-08	GAAGAGTGGTACTTTGGAA	7
HCK						
HCK	NM_002110	4504356	3055	D-003141-05	GAGATACCGTGAAACATTA	7
HCK	NM_002110	4504356	3055	D-003141-06	GCAGGGAGATACCGTGAAA	7
HCK	NM_002110	4504356	3055	D-003141-07	CATCGTGGTTGCCCTGTAT	7
HCK	NM_002110	4504356	3055	D-003141-08	TGTGTAAGATTGCTGACTT	7
ITK						
ITK	NM_005546	21614549	3702	D-003144-05	CAAATAATCTGGAAACCTA	7
ITK	NM_005546	21614549	3702	D-003144-06	GAAGAAACGAGGAATAATA	7
ITK	NM_005546	21614549	3702	D-003144-07	GAAACTCTCTCATCCCAA	7

ITK	NM_005546	21614549	3702	D-003144-08	GGAATGGGCATGAAGGATA	7
JAK1						
JAK1	NM_002227	4504802	3716	D-003145-05	CCACATAGCTGATCTGAAA	7
JAK1	NM_002227	4504802	3716	D-003145-06	TGAAATCACTCACATTGTA	7
JAK1	NM_002227	4504802	3716	D-003145-07	TAAGGAACCTCTATCATGA	7
JAK1	NM_002227	4504802	3716	D-003145-08	GCAGGTGGCTGTAAATCT	7
JAK2						
JAK2	NM_004972	13325062	3717	D-003146-05	GCAAATAGATCCAGTTCTT	7
JAK2	NM_004972	13325062	3717	D-003146-06	GAGCAAAGATCCAAGACTA	7
JAK2	NM_004972	13325062	3717	D-003146-07	GCCAGAAACTTGAAACTTA	7
JAK2	NM_004972	13325062	3717	D-003146-08	GTACAGATTTTCGAGATTT	7
JAK3						
JAK3	NM_000215	4557680	3718	D-003147-05	GCGCCTATCTTTCTCCTTT	7
JAK3	NM_000215	4557680	3718	D-003147-06	CCAGAAATCGTAGACATTA	7
JAK3	NM_000215	4557680	3718	D-003147-07	CCTCATCTCTTCAGACTAT	7
JAK3	NM_000215	4557680	3718	D-003147-08	TGTACGAGCTCTTCACCTA	7
KDR						
KDR	NM_002253	11321596	3791	D-003148-05	GGAAATCTCTTGCAAGCTA	7
KDR	NM_002253	11321596	3791	D-003148-06	GATTACAGATCTCCATTTA	7
KDR	NM_002253	11321596	3791	D-003148-07	GCAGACAGATCTACGTTTG	7
KDR	NM_002253	11321596	3791	D-003148-08	GCGATGGCCTCTTCTGTAA	7
KIAA1079						
KIAA1079	NM_014916	7662475	22853	D-003149-05	GAAATTCTCTCAACTGATG	7
KIAA1079	NM_014916	7662475	22853	D-003149-06	GCAGAGGTCTTCACACTTT	7
KIAA1079	NM_014916	7662475	22853	D-003149-07	TAAATGATCTTCAGACAGA	7
KIAA1079	NM_014916	7662475	22853	D-003149-08	GAGCAGCCCTACTCTGATA	7
KIT						
KIT	NM_000222	4557694	3815	D-003150-05	AAACACGGCTTAAGCAATT	7
KIT	NM_000222	4557694	3815	D-003150-06	GAACAGAACCTTCACTGAT	7
KIT	NM_000222	4557694	3815	D-003150-07	GGGAAGCCCTCATGTCTGA	7
KIT	NM_000222	4557694	3815	D-003150-08	GCAATTCCATTTATGTGTT	7
LCK						
LCK	NM_005356	20428651	3932	D-003151-05	GAACTGCCATTATCCCATA	7
LCK	NM_005356	20428651	3932	D-003151-06	GAGAGGTGGTGAAACATTA	7
LCK	NM_005356	20428651	3932	D-003151-07	GGGCCAAGTTTCCCATTAA	7
LCK	NM_005356	20428651	3932	D-003151-08	GCACGCTGCTCATCCGAAA	7
LTK						
LTK	NM_002344	4505044	4058	D-003152-05	TGAATTCACCTCCTGCCAAT	7
LTK	NM_002344	4505044	4058	D-003152-06	GTGGCAACCTCAACACTGA	7
LTK	NM_002344	4505044	4058	D-003152-07	GGAGCTAGCTGTGGATAAC	7
LTK	NM_002344	4505044	4058	D-003152-08	GCAAGTTTCGCCATCAGAA	7
LYN						
LYN	NM_002350	4505054	4067	D-003153-05	GCAGATGGCTTGTGCAGAA	7
LYN	NM_002350	4505054	4067	D-003153-06	GGAGAAGGCTTGTATTAGT	7
LYN	NM_002350	4505054	4067	D-003153-07	GATGAGCTCTATGACATTA	7
LYN	NM_002350	4505054	4067	D-003153-08	GGTGCTAAGTTCCCTATTA	7
MATK						
MATK	NM_002378	21450841	4145	D-003154-05	TGAAGAATATCAAGTGTGA	7
MATK	NM_002378	21450841	4145	D-003154-06	CCGCTCAGCTCCTGCAGTT	7
MATK	NM_002378	21450841	4145	D-003154-07	TACTGAACCTGCAGCATTT	7
MATK	NM_002378	21450841	4145	D-003154-08	TGGGAGGTCTTCTCATATG	8
MERTK						
MERTK	NM_006343	5453737	10461	D-003155-05	GAACTTACCTTACATAGCT	8
MERTK	NM_006343	5453737	10461	D-003155-06	GGACCTGCATACTTACTTA	8
MERTK	NM_006343	5453737	10461	D-003155-07	TGACAGGAATCTTCTAATT	8
MERTK	NM_006343	5453737	10461	D-003155-08	GGTAATGGCTCAGTCATGA	8

MET						
MET	NM 000245	4557746	4233	D-003156-05	GAAAGAACCTCTCAACATT	8
MET	NM 000245	4557746	4233	D-003156-06	GGACAAGGCTGACCATATG	8
MET	NM 000245	4557746	4233	D-003156-07	CCAATGACCTGCTGAAATT	8
MET	NM 000245	4557746	4233	D-003156-08	GAGCATACATTAAACCAA	8
MST1R						
MST1R	NM 002447	4505264	4486	D-003157-05	GGATGGAGCTGCTGGCTTT	8
MST1R	NM 002447	4505264	4486	D-003157-06	CTGCAGACCTATAGATTTA	8
MST1R	NM 002447	4505264	4486	D-003157-07	GCACCTGTCTCACTCTTGA	8
MST1R	NM 002447	4505264	4486	D-003157-08	GAAAGAGTCCATCCAGCTA	8
MUSK						
MUSK	NM 005592	5031926	4593	D-003158-05	GAAGAAGCCTCGGCAGATA	8
MUSK	NM 005592	5031926	4593	D-003158-06	GTAATAATCTCCATCATGT	8
MUSK	NM 005592	5031926	4593	D-003158-07	GGAATGAACTGAAAGTAGT	8
MUSK	NM 005592	5031926	4593	D-003158-08	GAGATTTCTGACTAGAA	8
NTRK1						
NTRK1	NM 002529	4585711	4914	D-003159-05	GGACAACCCTTTTCGAGTTC	8
NTRK1	NM 002529	4585711	4914	D-003159-06	CCAGTGACCTCAACAGGAA	8
NTRK1	NM 002529	4585711	4914	D-003159-07	CCACAATACTTCAGTGATG	8
NTRK1	NM 002529	4585711	4914	D-003159-08	GAAGAGTGGTCTCCGTTTC	8
NTRK2						
NTRK2	NM 006180	21361305	4915	D-003160-05	GAACAGAAGTAATGAAATC	8
NTRK2	NM 006180	21361305	4915	D-003160-06	GTAATGCTGTTTCTGCTTA	8
NTRK2	NM 006180	21361305	4915	D-003160-07	GCAAGACACTCCAAGTTTG	8
NTRK2	NM 006180	21361305	4915	D-003160-08	GAAAGTCTATCACATTATC	8
NTRK3						
NTRK3	NM 002530	4505474	4916	D-003161-05	GAGCGAATCTGCTAGTGAA	8
NTRK3	NM 002530	4505474	4916	D-003161-06	GAAGTTCACTACAGAGAGT	8
NTRK3	NM 002530	4505474	4916	D-003161-07	GGTCGACGGTCCAAATTTG	8
NTRK3	NM 002530	4505474	4916	D-003161-08	GAATATCACTTCCATACAC	8
PDGFRA						
PDGFRA	NM 006206	15451787	5156	D-003162-05	GAAACTTCCTGGACTATTT	8
PDGFRA	NM 006206	15451787	5156	D-003162-06	GAGATTTGGTCAACTATTT	8
PDGFRA	NM 006206	15451787	5156	D-003162-07	GCACGCCGCTTCCTGATAT	8
PDGFRA	NM 006206	15451787	5156	D-003162-08	CATCAGAGCTGGATCTAGA	8
PDGFRB						
PDGFRB	NM 002609	15451788	5159	D-003163-05	GAAAGGAGACGTCAAATAT	8
PDGFRB	NM 002609	15451788	5159	D-003163-06	GGAATGAGGTGGTCAACTT	8
PDGFRB	NM 002609	15451788	5159	D-003163-07	CAACGAGTCTCCAGTGCTA	8
PDGFRB	NM 002609	15451788	5159	D-003163-08	GAGAGGACCTGCCGAGCAA	8
PTK2						
PTK2	NM 005607	27886592	5747	D-003164-05	GAAGTTGGGTTGTCTAGAA	8
PTK2	NM 005607	27886592	5747	D-003164-06	GAAGAACAATGATGTAATC	8
PTK2	NM 005607	27886592	5747	D-003164-07	GGAAATTGCTTTGAAGTTG	8
PTK2	NM 005607	27886592	5747	D-003164-08	GTTCAAGCTGGATTATTT	8
PTK2B						
PTK2B	NM 004103	27886583	2185	D-003165-05	GAACATGGCTGACCTCATA	8
PTK2B	NM 004103	27886583	2185	D-003165-06	GGACCACGCTGCTCTATTT	8
PTK2B	NM 004103	27886583	2185	D-003165-07	GGACGAGGACTATTACAAA	8
PTK2B	NM 004103	27886583	2185	D-003165-08	TGGCAGAGCTCATCAACAA	8
PTK6						
PTK6	NM 005975	27886594	5753	D-003166-05	GAGAAAGTCCTGCCCGTTT	8
PTK6	NM 005975	27886594	5753	D-003166-06	TGAAGAAGCTGCGGCACAA	8
PTK6	NM 005975	27886594	5753	D-003166-07	CCGCGACTCTGATGAGAAA	8
PTK6	NM 005975	27886594	5753	D-003166-08	TGCCCGAGCTTGTGAACATA	8
PTK7						

1) Harm 2110

PTK7	NM 002821	27886610	5754	D-003167-05	GAGAGAAGCCCACTATTAA	8
PTK7	NM 002821	27886610	5754	D-003167-06	CGAGAGAAGCCCACTATTA	8
PTK7	NM 002821	27886610	5754	D-003167-07	GGAGGGAGTTGGAGATGTT	8
PTK7	NM 002821	27886610	5754	D-003167-08	GAAGACATGCCGCTATTTG	8
PTK9						
PTK9	NM 002822	4506274	5756	D-003168-05	GAAGAACTACGACAGATTA	8
PTK9	NM 002822	4506274	5756	D-003168-09	GAAGGAGACTATTTAGAGT	8
PTK9	NM 002822	4506274	5756	D-003168-10	GAGCGGATGCTGTATTCTA	8
PTK9	NM 002822	4506274	5756	D-003168-11	CTGCAGACTTCCTTTATGA	8
PTK9L						
PTK9L	NM 007284	31543446	11344	D-003169-05	AGAGAGAGCTCCAGCAGAT	8
PTK9L	NM 007284	31543446	11344	D-003169-06	TTAACGAGGTGAAGACAGA	8
PTK9L	NM 007284	31543446	11344	D-003169-07	ACACAGAGCCCACGGATGT	8
PTK9L	NM 007284	31543446	11344	D-003169-08	GCTGGGATCAGGACTATGA	8
RET						
RET	NM 000323	21536316	5979	D-003170-05	GCAAAGACCTGGAGAAGAT	8
RET	NM 000323	21536316	5979	D-003170-06	GCACACGGCTGCATGAGAA	8
RET	NM 000323	21536316	5979	D-003170-07	GAAGTGGCTGGAGAGAGT	8
RET	NM 000323	21536316	5979	D-003170-08	TTAAATGGATGGCAATTGA	8
ROR1						
ROR1	NM 005012	4826867	4919	D-003171-05	GCAAGCATCTTTACTAGGA	8
ROR1	NM 005012	4826867	4919	D-003171-06	GAGCAAGGCTAAAGAGCTA	8
ROR1	NM 005012	4826867	4919	D-003171-07	GAGAGCAACTTCATGTAAA	8
ROR1	NM 005012	4826867	4919	D-003171-08	GAGAATGTCCTGTGTCAA	8
ROR2						
ROR2	NM 004560	19743897	4920	D-003172-05	GGAAGTCTGCTGCTGCCTAT	8
ROR2	NM 004560	19743897	4920	D-003172-06	GCAGGTGCCTCCTCAGATG	8
ROR2	NM 004560	19743897	4920	D-003172-07	GCAATGTGCTAGTGTACGA	8
ROR2	NM 004560	19743897	4920	D-003172-08	GAAGACAGAATATGGTTCA	8
ROS1						
ROS1	NM 002944	19924164	6098	D-003173-05	GAGGAGACCTTCTTACTTA	8
ROS1	NM 002944	19924164	6098	D-003173-06	TTACAGAGGTTTCAGGATTA	8
ROS1	NM 002944	19924164	6098	D-003173-07	GAACAAACCTAAGCATGAA	8
ROS1	NM 002944	19924164	6098	D-003173-08	GAAAGAGCACTTCAAATAA	8
RYK						
RYK	NM 002958	11863158	6259	D-003174-05	GAAAGATGGTTACCGAATA	8
RYK	NM 002958	11863158	6259	D-003174-06	CAAAGTAGATTCTGAAGTT	8
RYK	NM 002958	11863158	6259	D-003174-07	TCACTACGCTCTATCCTTT	8
RYK	NM 002958	11863158	6259	D-003174-08	GGTGAAGGATATAGCAATA	8
SRC						
SRC	NM 005417	21361210	6714	D-003175-05	GAGAACCTGGTGTGCAAAG	8
SRC	NM 005417	21361210	6714	D-003175-09	GAGAGAACCTGGTGTGCAA	8
SRC	NM 005417	21361210	6714	D-003175-10	GGAGTTTGCTGGACTTTCT	8
SRC	NM 005417	21361210	6714	D-003175-11	GAAAGTGAGACCACGAAAG	8
SYK						
SYK	NM 003177	21361552	6850	D-003176-05	GGAATAATCTCAAGAATCA	8
SYK	NM 003177	21361552	6850	D-003176-06	GAAGTGGGCTCTGGTAATT	8
SYK	NM 003177	21361552	6850	D-003176-07	GGAAGAATCTGAGCAAATT	8
SYK	NM 003177	21361552	6850	D-003176-08	GAACAGACATGTCAAGGAT	8
TEC						
TEC	NM 003215	4507428	7006	D-003177-05	GAAATTGTCTAGTAAGTGA	8
TEC	NM 003215	4507428	7006	D-003177-06	CACCTGAAGTGTTTAATTA	8
TEC	NM 003215	4507428	7006	D-003177-07	GTACAAAGTCGCAATCAAA	8
TEC	NM 003215	4507428	7006	D-003177-08	TGGAGGAGATTCTTATTAA	8
TEK						
TEK	NM 000459	4557868	7010	D-003178-05	GAAAGAATATGCCTCCAAA	8

TEK	NM 000459	4557868	7010	D-003178-06	GGAATGACATCAAATTTCA	8
TEK	NM 000459	4557868	7010	D-003178-07	TGAAGTACCTGATATTCTA	8
TEK	NM 000459	4557868	7010	D-003178-08	CGAAAGACCTACGTGAATA	8
TIE						
TIE	NM 005424	4885630	7075	D-003179-05	GAGAGGAGGTTTATGTGAA	8
TIE	NM 005424	4885630	7075	D-003179-06	GGGACAGCCTCTACCCTTA	8
TIE	NM 005424	4885630	7075	D-003179-07	GAAGTTCTGTGCAAATTGG	8
TIE	NM 005424	4885630	7075	D-003179-08	CAACATGGCCTCAGAACTG	9
TNK1						
TNK1	NM 003985	4507610	8711	D-003180-05	GTTCTGGGCCTAAGTCTAA	9
TNK1	NM 003985	4507610	8711	D-003180-06	GAAGTGGGTCTACAAGATC	9
TNK1	NM 003985	4507610	8711	D-003180-07	CGAGAGGTATCGGTCATGA	9
TNK1	NM 003985	4507610	8711	D-003180-08	GGCGCATCCTGGAGCATTA	9
TXK						
TXK	NM 003328	4507742	7294	D-003181-05	GAACATCTATTGAGACAAG	9
TXK	NM 003328	4507742	7294	D-003181-06	TCAAGGCACTTTATGATTT	9
TXK	NM 003328	4507742	7294	D-003181-07	GGAGAGGAATGGCTATATT	9
TXK	NM 003328	4507742	7294	D-003181-08	GGATATATGTGAAGGAATG	9
TYK2						
TYK2	NM 003331	4507748	7297	D-003182-05	GAGGAGATCCACCACTTTA	9
TYK2	NM 003331	4507748	7297	D-003182-06	GCATCCACATTGCACATAA	9
TYK2	NM 003331	4507748	7297	D-003182-07	TCAAATACCTAGCCACACT	9
TYK2	NM 003331	4507748	7297	D-003182-08	CAATCTTGCTGACGTCTTG	9
TYRO3						
TYRO3	NM 006293	27597077	7301	D-003183-05	GGTAGAAGGTGTGCCATTT	9
TYRO3	NM 006293	27597077	7301	D-003183-06	ACGCTGAGATTTACAATA	9
TYRO3	NM 006293	27597077	7301	D-003183-07	GGATGGCTCCTTTGTGAAA	9
TYRO3	NM 006293	27597077	7301	D-003183-08	GAGAGGAACCTACGAAGATC	9
YES1						
YES1	NM 005433	21071041	7525	D-003184-05	GAAGGACCCTGATGAAAGA	9
YES1	NM 005433	21071041	7525	D-003184-06	TAAGAAGGGTGAAAGATTT	9
YES1	NM 005433	21071041	7525	D-003184-07	TCAAGAAGCTCAGATAATG	9
YES1	NM 005433	21071041	7525	D-003184-08	CAGAATCCCTCCATGAATT	9

Table VIII

5

Gene Name	Acc#	GI	Locus Link	Duplex #	Full Sequence	SEQ. ID NO.
APC2						
APC2	NM 013366	7549800	29882	D-003200-05	GCAAGGACCTCTTCATCAA	921
APC2	NM 013366	7549800	29882	D-003200-06	GAGAAGAAGTCCACACTAT	922
APC2	NM 013366	7549800	29882	D-003200-07	GGAATGCCATCTCCCAATG	923
APC2	NM 013366	7549800	29882	D-003200-09	CAACACGTGTGACATCATC	924
ATM						
ATM	NM 000051	20336202	472	D-003201-05	GCAAGCAGCTGAAACAAAT	925
ATM	NM 000051	20336202	472	D-003201-06	GAATGTTGCTTTCTGAATT	926
ATM	NM 000051	20336202	472	D-003201-07	GACCTGAAGTCTTATTTAA	927
ATM	NM 000051	20336202	472	D-003201-08	AGACAGAATTCCCAAATAA	928
ATR						
ATR	NM 001184	20143978	545	D-003202-05	GAACAACACTGCTGGTTTG	929
ATR	NM 001184	20143978	545	D-003202-06	GAAGTCATCTGTTTATTAT	930
ATR	NM 001184	20143978	545	D-003202-07	GAAATAAGGTAGACTCAAT	931
ATR	NM 001184	20143978	545	D-003202-08	CAACATAAATCCAAGAAGA	932
BTAK						

BTAK	NM 003600	3213196	6790	D-003545-04	CAAAGAATCAGCTAGCAAA	933
BTAK	NM 003600	3213196	6790	D-003203-05	GAAGAGAGTTATTCATAGA	934
BTAK	NM 003600	3213196	6790	D-003203-07	CAAATGCCCTGTCTTACTG	935
STK6	NM 003600	3213196	6790	D-003203-09	TCTCGTGA CT CAGCAAATT	936
CCNA1						
CCNA1	NM 003914	16306528	8900	D-003204-05	GAACCTGGCTAAGTACGTA	937
CCNA1	NM 003914	16306528	8900	D-003204-06	GCAGATCCATTCTTGAAAT	938
CCNA1	NM 003914	16306528	8900	D-003204-07	TCACAAGAATCAGGTGTTA	939
CCNA1	NM 003914	16306528	8900	D-003204-08	CATAAAGCGTACCTTGATA	940
CCNA2						
CCNA2	NM 001237	16950653	890	D-003205-05	GCTGTGAACTACATTGATA	941
CCNA2	NM 001237	16950653	890	D-003205-06	GATGATACCTACACCAAGA	942
CCNA2	NM 001237	16950653	890	D-003205-07	GCTGTTAGCCTCAAAGTTT	943
CCNA2	NM 001237	16950653	890	D-003205-08	AAGCTGGCCTGAATCATT	944
CCNB1						
CCNB1	NM 031966	14327895	891	D-003206-05	CAACATTACCTGTCATATA	945
CCNB1	NM 031966	14327895	891	D-003206-06	CCAAATACCTGATGGAAC	946
CCNB1	NM 031966	14327895	891	D-003206-07	GAAATGTACCCTCCAGAAA	947
CCNB1	NM 031966	14327895	891	D-003206-08	GCACCTGGCTAAGAATGTA	948
CCNB2						
CCNB2	NM 004701	10938017	9133	D-003207-05	CAACAAATGTCAACAAACA	949
CCNB2	NM 004701	10938017	9133	D-003207-06	GCAGCAAACCTCCTGAAGAT	950
CCNB2	NM 004701	10938017	9133	D-003207-07	CCAGTGATTTGGAGAATAT	951
CCNB2	NM 004701	10938017	9133	D-003207-08	GTGACTACGTAAAGGATAT	952
CCNB3						
CCNB3	NM 033031	14719419	85417	D-003208-05	TGAACAAACTGCTGACTTT	953
CCNB3	NM 033031	14719419	85417	D-003208-06	GCTAGCTGCTGCCTCCTTA	954
CCNB3	NM 033031	14719419	85417	D-003208-07	CAACTCACCTCGTGTGGAT	955
CCNB3	NM 033031	14719419	85417	D-003208-08	GTGGATCTCTACCTAATGA	956
CCNC						
CCNC	NM 005190	7382485	892	D-003209-05	GCAGAGCTCCCACTATTTG	957
CCNC	NM 005190	7382485	892	D-003209-06	GGAGTAGTTTCAAATACAA	958
CCNC	NM 005190	7382485	892	D-003209-07	GACCTTTGCTCCAGTATGT	959
CCNC	NM 005190	7382485	892	D-003209-08	GAGATTCTATGCCAGGTAT	960
CCND1						
CCND1	NM 053056	16950654	595	D-003210-05	TGAACAAGCTCAAGTGGA	961
CCND1	NM 053056	16950654	595	D-003210-06	CCAGAGTGATCAAGTGTGA	962
CCND1	NM 053056	16950654	595	D-003210-07	GTTCGTGGCCTCTAAGATG	963
CCND1	NM 053056	16950654	595	D-003210-08	CCGAGAAGCTGTGCATCTA	964
CCND2						
CCND2	NM 001759	16950656	894	D-003211-06	TGAATTACCTGGACCGTTT	965
CCND2	NM 001759	16950656	894	D-003211-07	CGGAGAAGCTGTGCATTTA	966
CCND2	NM 001759	16950656	894	D-003211-08	CTACAGACGTGCGGGATAT	967
CCND2	NM 001759	16950656	894	D-003211-09	CAACACAGACGTGGATTGT	968
CCND3						
CCND3	NM 001760	16950657	896	D-003212-05	GGACCTGGCTGCTGTGATT	969
CCND3	NM 001760	16950657	896	D-003212-06	GATTATACCTTTGCCATGT	970
CCND3	NM 001760	16950657	896	D-003212-07	GACCAGCACTCCTACAGAT	971
CCND3	NM 001760	16950657	896	D-003212-08	TGCGGAAGATGCTGGCTTA	972
CCNE1						
CCNE1	NM 001238	17318558	898	D-003213-05	GTA CTGAGCTGGGCAAATA	973
CCNE1	NM 001238	17318558	898	D-003213-06	GGAAATCTATCCTCCAAAG	974
CCNE1	NM 001238	17318558	898	D-003213-07	GGAGGTGTGTGAAGTCTAT	975
CCNE1	NM 001238	17318558	898	D-003213-08	CTAAATGACTTACATGAAG	976
CCNE2						
CCNE2	NM 057749	17318564	9134	D-003214-05	GGATGGAACTCATTATATT	977

CCNE2	NM_057749	17318564	9134	D-003214-06	GCAGATATGTTTCATGACAA	978
CCNE2	NM_057749	17318564	9134	D-003214-07	CATAATATCCAGACACATA	979
CCNE2	NM_057749	17318564	9134	D-003214-08	TAAGAAAGCCTCAGGTTTG	980
CCNF						
CCNF	NM_001761	4502620	899	D-003215-05	TCACAAAGCATCCATATTG	981
CCNF	NM_001761	4502620	899	D-003215-06	GAAGTCATGTTTACAGTGT	982
CCNF	NM_001761	4502620	899	D-003215-07	TAGCCTACCTCTACAATGA	983
CCNF	NM_001761	4502620	899	D-003215-08	GCACCCGGTTTATCAGTAA	984
CCNG1						
CCNG1	NM_004060	8670528	900	D-003216-05	GATAATGGCCTCAGAATGA	985
CCNG1	NM_004060	8670528	900	D-003216-06	GCACGGCAATTGAAGCATA	986
CCNG1	NM_004060	8670528	900	D-003216-07	GGAATAGAATGTCTTCAGA	987
CCNG1	NM_004060	8670528	900	D-003216-08	TAACTCACCTTCCAACAAT	988
CCNG2						
CCNG2	NM_004354	4757935	901	D-003217-05	GGAGAGAGTTGGTTTCTAA	989
CCNG2	NM_004354	4757935	901	D-003217-06	GGTGAAACCTAAACATTTG	990
CCNG2	NM_004354	4757935	901	D-003217-07	GAAATACTGAGCCTTGATA	991
CCNG2	NM_004354	4757935	901	D-003217-08	TGCCAAAGTTGAAGATTTA	992
CCNH						
CCNH	NM_001239	17738313	902	D-003218-05	GCTGATGACTTTCTTAATA	993
CCNH	NM_001239	17738313	902	D-003218-06	CAACTTAATTTCCACCTTA	994
CCNH	NM_001239	17738313	902	D-003218-07	ATACACACCTTCCCAAATT	995
CCNH	NM_001239	17738313	902	D-003218-08	GCTATGAAGATGATGATTA	996
CCNI						
CCNI	NM_006835	17738314	10983	D-003219-05	GCAAGCAGACCTCTACTAA	997
CCNI	NM_006835	17738314	10983	D-003219-07	TGAGAGAATTCCAGTACTA	998
CCNI	NM_006835	17738314	10983	D-003219-08	GGAATCAAACGGCTCTATA	999
CCNI	NM_006835	17738314	10983	D-003219-09	GAATTGGGATCTTCACACA	1000
CCNT1						
CCNT1	NM_001240	17978465	904	D-003220-05	TATCAACACTGCTATAGTA	1001
CCNT1	NM_001240	17978465	904	D-003220-06	GAACAAACGTCCTGGTGAT	1002
CCNT1	NM_001240	17978465	904	D-003220-07	GCACAAGACTCACCCATCT	1003
CCNT1	NM_001240	17978465	904	D-003220-08	GCACAGACTTCTTACTTCA	1004
CCNT2A						
CCNT2A	NM_001241	17978467	905	D-003221-05	GCACAGACATCCTATTTCA	1005
CCNT2A	NM_001241	17978467	905	D-003221-06	GCAGGGACCTTCTATATCA	1006
CCNT2A	NM_001241	17978467	905	D-003221-07	GAACAGCTATATTCACAGA	1007
CCNT2A	NM_001241	17978467	905	D-003221-09	TTATATAGCTGCCCAGGTA	1008
CCNT2B						
CCNT2B	NM_058241	17978468	905	D-003222-05	GCACAGACATCCTATTTCA	1009
CCNT2B	NM_058241	17978468	905	D-003222-06	GCAGGGACCTTCTATATCA	1010
CCNT2B	NM_058241	17978468	905	D-003222-07	GAACAGCTATATTCACAGA	1011
CCNT2B	NM_058241	17978468	905	D-003222-08	GGTGAAATGTACCCAGTTA	1012
CDC16						
CDC16	NM_003903	14110370	8881	D-003223-05	GTAGATGGCTTGCAAGAGA	1013
CDC16	NM_003903	14110370	8881	D-003223-06	TAAAGTAGCTTCACTCTCT	1014
CDC16	NM_003903	14110370	8881	D-003223-07	GCTACAAGCTTACTTCTGT	1015
CDC16	NM_003903	14110370	8881	D-003223-08	TGGAAGAGCCCATCAATAA	1016
CDC2						
CDC2	NM_033379	27886643	983	D-003552-01	GTACAGATCTCCAGAAGTA	1017
CDC2	NM_033379	27886643	983	D-003552-02	GATCAACTCTTCAGGATTT	1018
CDC2	NM_033379	27886643	983	D-003552-03	GGTTATATCTCATCTTTGA	1019
CDC2	NM_033379	27886643	983	D-003552-04	GAACCTCGTCATCCAAATA	1020
CDC20						
CDC20	NM_001255	4557436	991	D-003225-05	GGAATATATATCCTCTGT	1021
CDC20	NM_001255	4557436	991	D-003225-06	GAAACGGCTTCGAAATATG	1022

CDC20	NM 001255	4557436	991	D-003225-07	GAAGACCTGCCGTTACATT	1023
CDC20	NM 001255	4557436	991	D-003225-08	CACCAGTGATCGACACATT	1024
CDC25A						
CDC25A	NM 001789	4502704	993	D-003226-05	GAAATTATGGCATCTGTTT	1025
CDC25A	NM 001789	4502704	993	D-003226-06	TACAAGGAGTTCTTTATGA	1026
CDC25A	NM 001789	4502704	993	D-003226-07	CCACGAGGACTTTAAAGAA	1027
CDC25A	NM 001789	4502704	993	D-003226-08	TGGGAAACATCAGGATTTA	1028
CDC25B						
CDC25B	NM 004358	11641416	994	D-003227-05	GCAGATACCCCTATGAATA	1029
CDC25B	NM 004358	11641416	994	D-003227-06	CTAGGTGCTTCTCTCTGA	1030
CDC25B	NM 004358	11641416	994	D-003227-07	GAGAGCTGATTGGAGATTA	1031
CDC25B	NM 004358	11641416	994	D-003227-08	AAAAGGACCTCGTCATGTA	1032
CDC25C						
CDC25C	NM 001790	12408659	995	D-003228-05	GAGCAGAAGTGGCCTATAT	1033
CDC25C	NM 001790	12408659	995	D-003228-06	CAGAAGAGATTTAGATGA	1034
CDC25C	NM 001790	12408659	995	D-003228-07	CCAGGGAGCCTTAACTTA	1035
CDC25C	NM 001790	12408659	995	D-003228-08	GAAACTTGGTGGACAGTGA	1036
CDC27						
CDC27	NM 001256	16554576	996	D-003229-06	CATGCAAGCTGAAAGAATA	1037
CDC27	NM 001256	16554576	996	D-003229-07	CAACACAAGTACCTAATCA	1038
CDC27	NM 001256	16554576	996	D-003229-08	GGAGATGGATCCTATTTAC	1039
CDC27	NM 001256	16554576	996	D-003229-09	GAAAAGCCATGATGATATT	1040
CDC34						
CDC34	NM 004359	16357476	997	D-003230-05	GCTCAGACCTCTTCTACGA	1041
CDC34	NM 004359	16357476	997	D-003230-06	GGACGAGGGCGATCTATAC	1042
CDC34	NM 004359	16357476	997	D-003230-07	GATCGGGAGTACACAGACA	1043
CDC34	NM 004359	16357476	997	D-003230-08	TGAACGAGCCCAACACCTT	1044
CDC37						
CDC37	NM 007065	16357478	11140	D-003231-05	GCGAGGAGACAGCCAATTA	1045
CDC37	NM 007065	16357478	11140	D-003231-06	CACAAGACCTTCGTGGAAA	1046
CDC37	NM 007065	16357478	11140	D-003231-07	ACAATCGTCATGCAATTTA	1047
CDC37	NM 007065	16357478	11140	D-003231-08	GAGGAGAAATGTGCACTCA	1048
CDC45L						
CDC45L	NM 003504	34335230	8318	D-003232-05	GCACACGGATCTCCTTTGA	1049
CDC45L	NM 003504	34335230	8318	D-003232-06	GCAAACACCTGCTCAAGTC	1050
CDC45L	NM 003504	34335230	8318	D-003232-07	TGAAGAGTCTGCAAATAAA	1051
CDC45L	NM 003504	34335230	8318	D-003232-08	GGACGTGGATGCTCTGTGT	1052
CDC6						
CDC6	NM 001254	16357469	990	D-003233-05	GAACACAGCTGTCCCAGAT	1053
CDC6	NM 001254	16357469	990	D-003233-06	GAGCAGAGATGTCCACTGA	1054
CDC6	NM 001254	16357469	990	D-003233-07	GGAAATATCTTAGCTACTG	1055
CDC6	NM 001254	16357469	990	D-003233-08	GGACGAAGATTGGTATTG	1056
CDC7						
CDC7	NM 003503	11038647	8317	D-003234-05	GGAATGAGGTACCTGATGA	1057
CDC7	NM 003503	11038647	8317	D-003234-06	CAGGAAAGGTGTTCAAAA	1058
CDC7	NM 003503	11038647	8317	D-003234-07	CTACACAAATGCACAAATT	1059
CDC7	NM 003503	11038647	8317	D-003234-08	GTACGGGAATATATGCTTA	1060
CDK10						
CDK10	NM 003674	32528262	8558	D-003235-05	GAACTGCTGTTGGGAACCA	1061
CDK10	NM 003674	32528262	8558	D-003235-06	GGAAGCAGCCCTACAACAA	1062
CDK10	NM 003674	32528262	8558	D-003235-07	GCACGCCAGTGAGAACAT	1063
CDK10	NM 003674	32528262	8558	D-003235-08	GGAAGCAGCCCTACAACAA	1064
CDK2						
CDK2	NM 001798	16936527	1017	D-003236-05	GAGCTTAACCATCCTAATA	1065
CDK2	NM 001798	16936527	1017	D-003236-06	GAGCTTAACCATCCTAATA	1066
CDK2	NM 001798	16936527	1017	D-003236-07	GTACCGAGCTCCTGAAATC	1067

CDK2	NM_001798	16936527	1017	D-003236-08	GAGAGGTGGTGGCGCTTAA	1068
CDK3						
CDK3	NM_001258	4557438	1018	D-003237-05	GAGCATTGGTTGCATCTTT	1069
CDK3	NM_001258	4557438	1018	D-003237-06	GATCGGAGAGGGCACCTAT	1070
CDK3	NM_001258	4557438	1018	D-003237-07	GAAGCTCTATCTGGTGTTT	1071
CDK3	NM_001258	4557438	1018	D-003237-08	GCAGAGATGGTGA CT CGAA	1072
CDK4						
CDK4	NM_000075	456426	1019	D-003238-05	GCAGCACTCTTATCTACAT	1073
CDK4	NM_000075	456426	1019	D-003238-06	GGAGGAGGCCTTCCCATCA	1074
CDK4	NM_000075	456426	1019	D-003238-07	TCGAAAGCCTCTCTTCTGT	1075
CDK4	NM_000075	456426	1019	D-003238-08	GTACCGAGCTCCCGAAGTT	1076
CDK5						
CDK5	NM_004935	4826674	1020	D-003239-05	TGACCAAGCTGCCAGACTA	1077
CDK5	NM_004935	4826674	1020	D-003239-06	GAGCTGAAATTGGCTGATT	1078
CDK5	NM_004935	4826674	1020	D-003239-07	CAACATCCCTGGTGAACGT	1079
CDK5	NM_004935	4826674	1020	D-003239-08	GGATTCCCGTCCGCTGTTA	1080
CDK6						
CDK6	NM_001259	16950658	1021	D-003240-05	GCAAAGACCTACTTCTGAA	1081
CDK6	NM_001259	16950658	1021	D-003240-06	GAAGAAGACTGGCCTAGAG	1082
CDK6	NM_001259	16950658	1021	D-003240-07	GGTCTGGACTTTCTTCATT	1083
CDK6	NM_001259	16950658	1021	D-003240-08	TAACAGATATCGATGA ACT	1084
CDK7						
CDK7	NM_001799	16950659	1022	D-003241-05	GGACATAGATCAGAAGCTA	1085
CDK7	NM_001799	16950659	1022	D-003241-06	CAATAGAGCTTATACACAT	1086
CDK7	NM_001799	16950659	1022	D-003241-07	CATACAAGGCTTATTCTTA	1087
CDK7	NM_001799	16950659	1022	D-003241-08	GGAGACGACTTACTAGATC	1088
CDK8						
CDK8	NM_001260	4502744	1024	D-003242-05	CCACAGTACTCACATCAGA	1089
CDK8	NM_001260	4502744	1024	D-003242-06	GCAATAACCACACTAATGG	1090
CDK8	NM_001260	4502744	1024	D-003242-07	GAAGAAAGTGAGAGTTGTT	1091
CDK8	NM_001260	4502744	1024	D-003242-08	GAACATGACCTCTGGCATA	1092
CDK9						
CDK9	NM_001261	17017983	1025	D-003243-05	GGCCAAACGTGGACA ACTA	1093
CDK9	NM_001261	17017983	1025	D-003243-06	TGACGTCCATGTTCCGAGTA	1094
CDK9	NM_001261	17017983	1025	D-003243-07	CCAACCAGACGGAGTTTGA	1095
CDK9	NM_001261	17017983	1025	D-003243-08	GAAGGTGGCTCTGAAGAAG	1096
CDKN1C						
CDKN1C	NM_000076	4557440	1028	D-003244-05	GACCAGAACCGCTGGGATT	1097
CDKN1C	NM_000076	4557440	1028	D-003244-06	GGACCGAAGTGGACAGCGA	1098
CDKN1C	NM_000076	4557440	1028	D-003244-08	GCAAGAGATCAGCGCCTGA	1099
CDKN1C	NM_000076	4557440	1028	D-003244-09	CCGCTGGGATTACGACTTC	1100
CDKN2B						
CDKN2B	NM_004936	17981693	1030	D-003245-05	GCGAGGAGAACAAGGGCAT	1101
CDKN2B	NM_004936	17981693	1030	D-003245-06	CCAACGGAGTCAACCGTTT	1102
CDKN2B	NM_004936	17981693	1030	D-003245-07	CGATCCAGGTCATGATGAT	1103
CDKN2B	NM_004936	17981693	1030	D-003245-08	CCTGGAAGCCGGCGCGGAT	1104
CDKN2C						
CDKN2C	NM_001262	17981697	1031	D-003246-05	GGACACCGCCTGTGATTTG	1105
CDKN2C	NM_001262	17981697	1031	D-003246-06	GCCAGGAGACTGCTACTTA	1106
CDKN2C	NM_001262	17981697	1031	D-003246-07	TGAAAGACCGAACTGGTTT	1107
CDKN2C	NM_001262	17981697	1031	D-003246-08	GAACCTGCCCTTGCACTTG	1108
CDKN2D						
CDKN2D	NM_001800	17981700	1032	D-003247-05	TGGCAGTTCAAGAGGGTCA	1109
CDKN2D	NM_001800	17981700	1032	D-003247-06	CTCAGGACCTCGTGGACAT	1110
CDKN2D	NM_001800	17981700	1032	D-003247-07	TGAAGGTCCTAGTGGAGCA	1111
CDKN2D	NM_001800	17981700	1032	D-003247-08	AGACGGCGCTGCAGGTCAT	1112

CDT1						
CDT1	NM_030928	19923847	81620	D-003248-05	CCAAGGAGGCACAGAAGCA	1113
CDT1	NM_030928	19923847	81620	D-003248-06	GCTTCAACGTGGATGAAGT	1114
CDT1	NM_030928	19923847	81620	D-003248-07	TCTCCGGGCCAGAAGATAA	1115
CDT1	NM_030928	19923847	81620	D-003248-08	GCGCAATGTTGGCCAGATC	1116
CENPA						
CENPA	NM_001809	4585861	1058	D-003249-05	GCACACACCTCTTGATAAG	1117
CENPA	NM_001809	4585861	1058	D-003249-06	GCAAGAGAAATATGTGTTA	1118
CENPA	NM_001809	4585861	1058	D-003249-07	TTACATGCAGGCCGAGTTA	1119
CENPA	NM_001809	4585861	1058	D-003249-08	GAGACAAGGTTGGCTAAAG	1120
CENPB						
CENPB	NM_001810	26105977	1059	D-003250-05	GGACATAGCCGCCTGCTTT	1121
CENPB	NM_001810	26105977	1059	D-003250-06	GCACGATCCTGAAGAACAA	1122
CENPB	NM_001810	26105977	1059	D-003250-07	GGAGGAGGGTGATGTTGAT	1123
CENPB	NM_001810	26105977	1059	D-003250-08	CCGAATGGCTGCAGAGTCT	1124
CENPC1						
CENPC1	NM_001812	4502778	1060	D-003251-05	GCGAATAGATTATCAAGGA	1125
CENPC1	NM_001812	4502778	1060	D-003251-06	GAACAGAATCCATCACAAA	1126
CENPC1	NM_001812	4502778	1060	D-003251-07	CCATAAACCTCACCCAGTA	1127
CENPC1	NM_001812	4502778	1060	D-003251-08	CAAGAGAACACGTTTGAAA	1128
CENPE						
CENPE	NM_001813	4502780	1062	D-003252-05	GAAGACAGCTCAAATAATA	1129
CENPE	NM_001813	4502780	1062	D-003252-06	CAACAAAGCTACTAAATCA	1130
CENPE	NM_001813	4502780	1062	D-003252-07	GGAAAGAAGTGCTACCATA	1131
CENPE	NM_001813	4502780	1062	D-003252-08	GGAAAGAAATGACACAGTT	1132
CENPF						
CENPF	NM_016343	14670380	1063	D-003253-05	GCGAATATCTGAATTAGAA	1133
CENPF	NM_016343	14670380	1063	D-003253-06	GGAAATTAATGCATCCTTA	1134
CENPF	NM_016343	14670380	1063	D-003253-07	GAGCGAGGCTGGTGGTTTA	1135
CENPF	NM_016343	14670380	1063	D-003253-08	CAAGTCATCTTTCATCTAA	1136
CENPH						
CENPH	NM_022909	21264590	64946	D-003254-05	GAAAGAAGAGATTGCAATT	1137
CENPH	NM_022909	21264590	64946	D-003254-06	CAGAACAAATTATGCAAGA	1138
CENPH	NM_022909	21264590	64946	D-003254-07	CTAGTGTGCTCATGGATAA	1139
CENPH	NM_022909	21264590	64946	D-003254-08	GAAACACCTATTAGAGCTA	1140
CHEK1						
CHEK1	NM_001274	20127419	1111	D-003255-05	CAAATTGGATGCAGACAAA	1141
CHEK1	NM_001274	20127419	1111	D-003255-06	GCAACAGTATTTGCGTATA	1142
CHEK1	NM_001274	20127419	1111	D-003255-07	GGACTTCTCTCCAGTAAAC	1143
CHEK1	NM_001274	20127419	1111	D-003255-08	AAAGATAGATGGTACAACA	1144
CHEK2						
CHEK2	NM_007194	22209010	11200	D-003256-02	CTCTTACATTGCATACATA	1145
CHEK2	NM_007194	22209010	11200	D-003256-03	TAAACGCCTGAAAGAAGCT	1146
CHEK2	NM_007194	22209010	11200	D-003256-04	GCATAGGACTCAAGTGTCA	1147
CHEK2	NM_007194	22209010	11200	D-003256-05	GAAATTGCACTGTCACTAA	1148
CNK						
CNK	NM_004073	4758015	1263	D-003257-05	GCGAGAAGATCCTAAATGA	1149
CNK	NM_004073	4758015	1263	D-003257-07	GCAAGTGGGTTGACTACTC	1150
CNK	NM_004073	4758015	1263	D-003257-08	GCACATCCGTTGGCCATCA	1151
CNK	NM_004073	4758015	1263	D-003257-09	GACCTCAAGTTGGGAAATT	1152
CRI1						
CRI1	NM_014335	7656937	23741	D-003258-05	GTGATGAGATTATTGATAG	1153
CRI1	NM_014335	7656937	23741	D-003258-06	GGACGAGGGCGAGGAATTT	1154
CRI1	NM_014335	7656937	23741	D-003258-07	GGAAACGGAGCCTTGCTAA	1155
CRI1	NM_014335	7656937	23741	D-003258-08	TCAATCGTCTGACCGAAGA	1156
E2F1						

E2F1	NM 005225	12669910	1869	D-003259-05	GAACAGGGCCACTGACTCT	1157
E2F1	NM 005225	12669910	1869	D-003259-06	TGGACCACCTGATGAATAT	1158
E2F1	NM 005225	12669910	1869	D-003259-07	CCCAGGAGGTCACTTCTGA	1159
E2F1	NM 005225	12669910	1869	D-003259-08	GGCTGGACCTGGAACTGA	1160
E2F2						
E2F2	NM 004091	34485718	1870	D-003260-05	GGGAGAAGACTCGGTATGA	1161
E2F2	NM 004091	34485718	1870	D-003260-06	GAGGACAACCTGCAGATAT	1162
E2F2	NM 004091	34485718	1870	D-003260-07	TGAAGGAGCTGATGAACAC	1163
E2F2	NM 004091	34485718	1870	D-003260-08	CCAAGAAGTTCATTACCT	1164
E2F3						
E2F3	NM 001949	12669913	1871	D-003261-05	GAAATTAGATGAACTGATC	1165
E2F3	NM 001949	12669913	1871	D-003261-06	TGAAGTGCCTGACTCAATA	1166
E2F3	NM 001949	12669913	1871	D-003261-07	GAACAAGGCAGCAGAAGTG	1167
E2F3	NM 001949	12669913	1871	D-003261-08	GAAACACACAGTCCAATGA	1168
E2F4						
E2F4	NM 001950	12669914	1874	D-003262-05	GGAGATTGCTGACAACTG	1169
E2F4	NM 001950	12669914	1874	D-003262-06	GAAGGTATCGGGCTAATCG	1170
E2F4	NM 001950	12669914	1874	D-003262-07	GTGCAGAAAGTCCAGGGAAT	1171
E2F4	NM 001950	12669914	1874	D-003262-08	GGACAGTGGTGAGCTCAGT	1172
E2F5						
E2F5	NM 001951	12669916	1875	D-003263-05	GCAGATGACTACAACITTA	1173
E2F5	NM 001951	12669916	1875	D-003263-06	GACATCAGCTACAGATATA	1174
E2F5	NM 001951	12669916	1875	D-003263-07	CAACATGTCTCTGAAAGAA	1175
E2F5	NM 001951	12669916	1875	D-003263-08	GAAGACATCTGTAATTGCT	1176
E2F6						
E2F6	NM 001952	12669917	1876	D-003264-05	TAAACAAGGTTGCAACGAA	1177
E2F6	NM 001952	12669917	1876	D-003264-06	TAGCATATGTGACCTATCA	1178
E2F6	NM 001952	12669917	1876	D-003264-07	GAAACCAGATTGGATGTTT	1179
E2F6	NM 001952	12669917	1876	D-003264-09	GGAACITTTCTGACTTATCA	1180
FOS						
FOS	NM 005252	6552332	2353	D-003265-05	GGGATAGCCTCTCTTACTA	1181
FOS	NM 005252	6552332	2353	D-003265-06	GAACAGTTATCTCCAGAAG	1182
FOS	NM 005252	6552332	2353	D-003265-07	GGAGACAGACCAACTAGAA	1183
FOS	NM 005252	6552332	2353	D-003265-08	AGACCGAGCCCTTTGATGA	1184
HIPK2						
HIPK2	NM 022740	13430859	28996	D-003266-06	GAGAATCACTCCAATCGAA	1185
HIPK2	NM 022740	13430859	28996	D-003266-07	AGACAGGGATTAAGTCAAA	1186
HIPK2	NM 022740	13430859	28996	D-003266-08	GGACAAAGACAAGTAGGTT	1187
HIPK2	NM 022740	13430859	28996	D-003266-09	GCACACACGTCAAATCATG	1188
HUS1						
HUS1	NM 004507	31077213	3364	D-003267-05	ACAAAGGCCTTATGCAATA	1189
HUS1	NM 004507	31077213	3364	D-003267-06	GAAGTGCACATAGATATTA	1190
HUS1	NM 004507	31077213	3364	D-003267-07	AAGCTTAACCTCATCCTTT	1191
HUS1	NM 004507	31077213	3364	D-003267-08	GAACTTCTTCAACGAATTT	1192
JUN						
JUN	NM 002228	7710122	3725	D-003268-05	TGGAAACGACCTTCTATGA	1193
JUN	NM 002228	7710122	3725	D-003268-06	GAACTGCACAGCCAGAACA	1194
JUN	NM 002228	7710122	3725	D-003268-07	GAGCTGGAGCGCCTGATAA	1195
JUN	NM 002228	7710122	3725	D-003268-08	TAACGCAGCAGTTGCAAAC	1196
JUNB						
JUNB	NM 002229	4504808	3726	D-003269-05	GCATCAAAGTGAGCGCAA	1197
JUNB	NM 002229	4504808	3726	D-003269-06	TGGAAGACCAAGAGCGCAT	1198
JUNB	NM 002229	4504808	3726	D-003269-07	CATACACAGCTACGGGATA	1199
JUNB	NM 002229	4504808	3726	D-003269-08	CCATCAACATGGAAGACCA	1200
LOC510						
53						

LOC510 53	NM_015895	20127542	51053	D-003270-05	GGAGAAAGGCGCTGTATGA	1201
LOC510 53	NM_015895	20127542	51053	D-003270-06	GAATAGTTCTGTCCCAAGA	1202
LOC510 53	NM_015895	20127542	51053	D-003270-07	GAACATGTACAGTATATGG	1203
LOC510 53	NM_015895	20127542	51053	D-003270-08	GCAGAAACAAGAAGAAATC	1204
MAD2L1						
MAD2L1	NM_002358	6466452	4085	D-003271-05	GAAAGATGGCAGTTTGATA	1205
MAD2L1	NM_002358	6466452	4085	D-003271-06	TAAATAATGTGGTGGAAACA	1206
MAD2L1	NM_002358	6466452	4085	D-003271-07	GAAATCCGTTCAAGTGATCA	1207
MAD2L1	NM_002358	6466452	4085	D-003271-08	TTACTCGAGTGCAGAAATA	1208
MAD2L2						
MAD2L2	NM_006341	6006019	10459	D-003272-05	GGAAGAGCGCGCTCATAAA	1209
MAD2L2	NM_006341	6006019	10459	D-003272-06	TGGAAGAGCGCGCTCATAA	1210
MAD2L2	NM_006341	6006019	10459	D-003272-07	AGCCACTCCTGGAGAAGAA	1211
MAD2L2	NM_006341	6006019	10459	D-003272-08	TGGAGAAATTCGTCTTTGA	1212
MCM2						
MCM2	NM_004526	33356546	4171	D-003273-05	GAAGATCTTTGCCAGCATT	1213
MCM2	NM_004526	33356546	4171	D-003273-06	GGATAAGGCTCGTCAGATC	1214
MCM2	NM_004526	33356546	4171	D-003273-07	CAGAGCAGGTGACATATCA	1215
MCM2	NM_004526	33356546	4171	D-003273-08	GCCGTGGGCTCCTGTATGA	1216
MCM3						
MCM3	NM_002388	33356548	4172	D-003274-05	GGACATCAATATTCTTCTA	1217
MCM3	NM_002388	33356548	4172	D-003274-06	GCCAGGACATCTCCAGTTA	1218
MCM3	NM_002388	33356548	4172	D-003274-07	GCAGGTATGACCAGTATAA	1219
MCM3	NM_002388	33356548	4172	D-003274-08	GGAAATGCCTCAAGTACAC	1220
MCM4						
MCM4	XM_030274	22047061	4173	D-003275-05	GGACATATCTATTCTTACT	1221
MCM4	XM_030274	22047061	4173	D-003275-06	GATGTTAGTTCACCACTGA	1222
MCM4	XM_030274	22047061	4173	D-003275-07	CCAGCTGCCTCATACTTTA	1223
MCM4	XM_030274	22047061	4173	D-003275-08	GAAAGTACAAGATCGGTAT	1224
MCM5						
MCM5	NM_006739	23510447	4174	D-003276-05	GAAGATCCCTGGCATCATC	1225
MCM5	NM_006739	23510447	4174	D-003276-06	GAACAGGGTTACCATCATG	1226
MCM5	NM_006739	23510447	4174	D-003276-07	GGACAACATTGACTTCATG	1227
MCM5	NM_006739	23510447	4174	D-003276-08	CCAAGGAGGTAGCTGATGA	1228
MCM6						
MCM6	NM_005915	33469920	4175	D-003277-05	GGAAAGAGCTCAGAGATGA	1229
MCM6	NM_005915	33469920	4175	D-003277-06	GAGCAGCGATGGAGAAATT	1230
MCM6	NM_005915	33469920	4175	D-003277-07	GGAAACACCTGATGTCAAT	1231
MCM6	NM_005915	33469920	4175	D-003277-08	CCAACATCTGCCGAAATC	1232
MCM7						
MCM7	NM_005916	33469967	4176	D-003278-05	GGAAATATCCCTCGTAGTA	1233
MCM7	NM_005916	33469967	4176	D-003278-06	GGAAGAAGCAGTTCAAGTA	1234
MCM7	NM_005916	33469967	4176	D-003278-07	CAACAAGCCTCGTGTGATC	1235
MCM7	NM_005916	33469967	4176	D-003278-08	GGAGAGAACACAAGGATTG	1236
MDM2						
MDM2	NM_002392	4505136	4193	D-003279-05	GGAGATATGTTGTGAAAGA	1237
MDM2	NM_002392	4505136	4193	D-003279-06	CCACAAATCTGATAGTATT	1238
MDM2	NM_002392	4505136	4193	D-003279-07	GATGAGGTATATCAAGTTA	1239
MDM2	NM_002392	4505136	4193	D-003279-08	GGAAGAAACCCAAGACAAA	1240
MKI67						
MKI67	NM_002417	19923216	4288	D-003280-05	GCACAAAGCTTGTTATAA	1241
MKI67	NM_002417	19923216	4288	D-003280-06	CCTAAGACCTGAACATTTT	1242
MKI67	NM_002417	19923216	4288	D-003280-07	CAAAGAGGAACACAAATTA	1243

MKI67	NM_002417	19923216	4288	D-003280-08	GTAAATGGGTCTGTTATTG	1244
MNAT1						
MNAT1	NM_002431	4505224	4331	D-003281-05	GGAAGAAGCTTTAGAAGTG	1245
MNAT1	NM_002431	4505224	4331	D-003281-06	TAGATGAGCTGGAGAGTTC	1246
MNAT1	NM_002431	4505224	4331	D-003281-07	GGACCTTGCTGGAGGCTAT	1247
MNAT1	NM_002431	4505224	4331	D-003281-08	GCAGATAGAGACATATGGA	1248
MYC						
MYC	NM_002467	31543215	4609	D-003282-05	CAGAGAAGCTGGCCTCCTA	1249
MYC	NM_002467	31543215	4609	D-003282-06	GAAACGACGAGAACAGTTG	1250
MYC	NM_002467	31543215	4609	D-003282-07	CGACGAGACCTTCATCAA	1251
MYC	NM_002467	31543215	4609	D-003282-08	CCACACATCAGCACAATA	1252
ORC1L						
ORC1L	NM_004153	31795543	4998	D-003283-05	GAACAGGAATTCCAAGACA	1253
ORC1L	NM_004153	31795543	4998	D-003283-06	TAAGAAACGTGCTCGAGTA	1254
ORC1L	NM_004153	31795543	4998	D-003283-07	GAGATCACCTCACCTTCTA	1255
ORC1L	NM_004153	31795543	4998	D-003283-08	GCAGAGAGCCCTTCTTGGA	1256
ORC2L						
ORC2L	NM_006190	32454751	4999	D-003284-05	GAAGAAACCTCCTATGAGA	1257
ORC2L	NM_006190	32454751	4999	D-003284-06	GAAGGGAACCTGATGGAGTA	1258
ORC2L	NM_006190	32454751	4999	D-003284-07	GAAGAATGATCCTGAGATT	1259
ORC2L	NM_006190	32454751	4999	D-003284-08	GAAGAGATGTTCAAGAATC	1260
ORC3L						
ORC3L	NM_012381	32483366	23595	D-003285-05	GGACTGCTGTGTAGATATA	1261
ORC3L	NM_012381	32483366	23595	D-003285-06	GAAGTATGACCATACTTG	1262
ORC3L	NM_012381	32483366	23595	D-003285-07	AAAGATCTCTCTGCCAATA	1263
ORC3L	NM_012381	32483366	23595	D-003285-08	CAGCACAGCTAAGAGAATA	1264
ORC4L						
ORC4L	NM_002552	32454749	5000	D-003286-06	GAAAGCACATTCCGTTTAT	1265
ORC4L	NM_002552	32454749	5000	D-003286-07	TGAAAGAACTCATGGAAAT	1266
ORC4L	NM_002552	32454749	5000	D-003286-08	GCTGAGAAGTGGAAATGAA	1267
ORC4L	NM_002552	32454749	5000	D-003286-09	CCAGTGATCTTCATATTAG	1268
ORC5L						
ORC5L	NM_002553	32454752	5001	D-003287-05	GAAATAACCTGTGAAACAT	1269
ORC5L	NM_002553	32454752	5001	D-003287-06	CAGATTACCTCTCTAGTGA	1270
ORC5L	NM_002553	32454752	5001	D-003287-07	GAAGTCCATATTACTCTA	1271
ORC5L	NM_002553	32454752	5001	D-003287-08	GTATTCAGCTGATTTCTAT	1272
ORC6L						
ORC6L	NM_014321	32454755	23594	D-003288-05	GAACATGGCTTCAAAGATA	1273
ORC6L	NM_014321	32454755	23594	D-003288-06	GGACAGGGCTTATTTAATT	1274
ORC6L	NM_014321	32454755	23594	D-003288-07	GAAAGAAGATAGTGGTTGA	1275
ORC6L	NM_014321	32454755	23594	D-003288-08	TATCAGAGCTGTCTTAAAT	1276
PCNA						
PCNA	NM_002592	33239449	5111	D-003289-05	GATCGAGGATGAAGAAGGA	1277
PCNA	NM_002592	33239449	5111	D-003289-07	GCCGAGATCTCAGCCATAT	1278
PCNA	NM_002592	33239449	5111	D-003289-09	GAGGCCTGCTGGGATATTA	1279
PCNA	NM_002592	33239449	5111	D-003289-10	GTGGAGAACTTGGAATGG	1280
PLK						
PLK	NM_005030	21359872	5347	D-003290-05	CAACCAAAGTCGAATATGA	1281
PLK	NM_005030	21359872	5347	D-003290-06	CAAGAAGAATGAATACAGT	1282
PLK	NM_005030	21359872	5347	D-003290-07	GAAGATGTCCATGGAAATA	1283
PLK	NM_005030	21359872	5347	D-003290-08	CAACACGCCTCATCTCTA	1284
PIN1						
PIN1	NM_006221	5453897	5300	D-003291-05	GGACCAAGGAGGAGGCCCT	1285
PIN1	NM_006221	5453897	5300	D-003291-06	CGTCCTGGCGGCAGGAGAA	1286
PIN1	NM_006221	5453897	5300	D-003291-07	CGGGAGAGGAGGACTTTGA	1287
PIN1	NM_006221	5453897	5300	D-003291-08	AGTCGGGAGAGGAGGACTT	1288

PIN1L						
PIN1L	NM_006222	5453899	5301	D-003292-06	CGACCTGGCGGCAGGAAAT	1289
PIN1L	NM_006222	5453899	5301	D-003292-07	AGGCAGGAGAGAAGGACTT	1290
PIN1L	NM_006222	5453899	5301	D-003292-08	GCTACATCCAGAAGATCAA	1291
PIN1L	NM_006222	5453899	5301	D-003292-09	GGACAGTGTTCACGGATTC	1292
RAD1						
RAD1	NM_002853	19718797	5810	D-003293-05	GAAGATGGACAAATATGTT	1293
RAD1	NM_002853	19718797	5810	D-003293-06	GGAAGAGTCTGTTACTTTT	1294
RAD1	NM_002853	19718797	5810	D-003293-07	GATAACAGAGGCTTCCTTT	1295
RAD1	NM_002853	19718797	5810	D-003293-08	GCATTAGTCCTATCTTGTA	1296
RAD17						
RAD17	NM_133338	19718783	5884	D-003294-05	GAATCAAGCTTCCATATGT	1297
RAD17	NM_133338	19718783	5884	D-003294-06	CAACAAAGCCCGAGGATAT	1298
RAD17	NM_133338	19718783	5884	D-003294-07	ACACATGCCTGGAGACTTA	1299
RAD17	NM_133338	19718783	5884	D-003294-08	CTACATAGATTCTTCATG	1300
RAD9A						
RAD9A	NM_004584	19924112	5883	D-003295-05	TCAGCAAACCTTGAATCTTA	1301
RAD9A	NM_004584	19924112	5883	D-003295-06	GACATTGACTCTTACATGA	1302
RAD9A	NM_004584	19924112	5883	D-003295-08	GGAAACCACTATAGGCAAT	1303
RAD9A	NM_004584	19924112	5883	D-003295-09	CGGACGACTTTGCCAATGA	1304
RB1						
RB1	NM_000321	19924112	5925	D-003296-05	GAAAGGACATGTGAACTTA	1305
RB1	NM_000321	19924112	5925	D-003296-06	GAAGAAGTATGATGTATTG	1306
RB1	NM_000321	4506434	5925	D-003296-07	GAAATGACTTCTACTCGAA	1307
RB1	NM_000321	4506434	5925	D-003296-08	GGAGGGAACATCTATATTT	1308
RBBP2						
RBBP2	NM_005056	4826967	5927	D-003297-05	CAAAGAAGCTGAATAAACT	1309
RBBP2	NM_005056	4826967	5927	D-003297-06	CAACACATATGGCGGATTT	1310
RBBP2	NM_005056	4826967	5927	D-003297-07	GGACAAACCTAGAAAGAAG	1311
RBBP2	NM_005056	4826967	5927	D-003297-08	GAAAGGCACTCTCTCTGTT	1312
RBL1						
RBL1	NM_002895	34577078	5933	D-003298-05	CAAGAGAAGTTGTGGCATA	1313
RBL1	NM_002895	34577078	5933	D-003298-06	CAGCAGCACTCCATTTATA	1314
RBL1	NM_002895	34577078	5933	D-003298-07	ACAGAAAGGTCTATCATTT	1315
RBL1	NM_002895	34577078	5933	D-003298-08	GGACATAAAGTTACAATTC	1316
RBL2						
RBL2	NM_005611	21361291	5934	D-003299-05	GAGCAGAGCTTAATCGAAT	1317
RBL2	NM_005611	21361291	5934	D-003299-06	GAGAATAGCCCTTGTGTGA	1318
RBL2	NM_005611	21361291	5934	D-003299-07	GGACTTAGTTTATGGAAAT	1319
RBL2	NM_005611	21361291	5934	D-003299-08	GAATTTAGATGAGCGGATA	1320
RBP1						
RBP1	NM_002899	8400726	5947	D-003300-05	GAGACAAGCTCCAGTGTGT	1321
RBP1	NM_002899	8400726	5947	D-003300-06	GCAAGCAAGTATTCAAGAA	1322
RBP1	NM_002899	8400726	5947	D-003300-07	GCAGGACGGTGACCATATG	1323
RBP1	NM_002899	8400726	5947	D-003300-08	GCAAGTGCATGACAACAGT	1324
RPA3						
RPA3	NM_002947	19923751	6119	D-003322-05	GGAAGTGGTTGGAAGAGTA	1325
RPA3	NM_002947	19923751	6119	D-003322-06	GAAGATAGCCATCCTTTTG	1326
RPA3	NM_002947	19923751	6119	D-003322-07	CATGCTAGCTCAATTCATC	1327
RPA3	NM_002947	19923751	6119	D-003322-08	GATCTTGGACTTTACAATG	1328
SKP1A						
SKP1A	NM_006930	25777710	6500	D-003323-05	GGAGAGATATTTGAAGTTG	1329
SKP1A	NM_006930	25777710	6500	D-003323-06	GGGAATGGATGATGAAGGA	1330
SKP1A	NM_006930	25777710	6500	D-003323-07	CAAACAATCTGTGACTATT	1331
SKP1A	NM_006930	25777710	6500	D-003323-08	TCAATTAAGTTGCAGAGTT	1332
SKP2						

SKP2	NM_005983	16306594	6502	D-003324-05	CATCTAGACTTAAGTGATA	1333
SKP2	NM_005983	16306594	6502	D-003324-06	GAAATCAGATCTCTCTACT	1334
SKP2	NM_005983	16306594	6502	D-003324-07	CTAAAGGTCTCTGGTGTTT	1335
SKP2	NM_005983	16306594	6502	D-003324-08	GATGGTACCCTTCAACTGT	1336
SNK						
SNK	NM_006622	5730054	10769	D-003325-05	GAAGACATCTACAAGCTTA	1337
SNK	NM_006622	5730054	10769	D-003325-06	GAAATACCTTCATGAACAA	1338
SNK	NM_006622	5730054	10769	D-003325-07	GAAGGTCAATGGCTCATAT	1339
SNK	NM_006622	5730054	10769	D-003325-08	CCGGAGATCTCGCGGATTA	1340
STK12						
STK12	NM_004217	4759177	9212	D-003326-07	CAGAAGAGCTGCACATTTG	1341
STK12	NM_004217	4759177	9212	D-003326-08	CCAAACTGCTCAGGCATAA	1342
STK12	NM_004217	4759177	9212	D-003326-09	ACGCGGCACTTCACAATTG	1343
STK12	NM_004217	4759177	9212	D-003326-10	TGGGACACCCGACATCTTA	1344
TFDP1						
TFDP1	NM_007111	34147667	7027	D-003327-05	GGAAGCAGCTCTTGCCAAA	1345
TFDP1	NM_007111	34147667	7027	D-003327-06	GAGGAGACTTGAAAGAATA	1346
TFDP1	NM_007111	34147667	7027	D-003327-07	GAAGGTAGAGGTGGAAAGA	1347
TFDP1	NM_007111	34147667	7027	D-003327-08	GCGAGAAGGTGCAGAGGAA	1348
TFDP2						
TFDP2	NM_006286	5454111	7029	D-003328-05	GAAAGTGTGTGAGAAAGTT	1349
TFDP2	NM_006286	5454111	7029	D-003328-06	CACAGGACCTTCTTGGTTA	1350
TFDP2	NM_006286	5454111	7029	D-003328-07	CGAAATCCCTGGTGCCAAA	1351
TFDP2	NM_006286	5454111	7029	D-003328-08	TGAGATCCATGATGACATA	1352
TP53						
TP53	NM_000546	8400737	7157	D-003329-05	GAGGTTGGCTCTGACTGTA	1353
TP53	NM_000546	8400737	7157	D-003329-06	CAGTCTACCTCCCGCCATA	1354
TP53	NM_000546	8400737	7157	D-003329-07	GCACAGAGGAAGAGAATCT	1355
TP53	NM_000546	8400737	7157	D-003329-08	GAAGAAACCACTGGATGGA	1356
TP63						
TP63	NM_003722	31543817	8626	D-003330-05	CATCATGTCTGGACTATTT	1357
TP63	NM_003722	31543817	8626	D-003330-06	CAAACAAGATTGAGATTAG	1358
TP63	NM_003722	31543817	8626	D-003330-07	GCACACAGACAAATGAATT	1359
TP63	NM_003722	31543817	8626	D-003330-08	CGACAGTCTTGTACAATTT	1360
TP73						
TP73	NM_005427	4885644	7161	D-003331-05	GCAAGCAGCCCATCAAGGA	1361
TP73	NM_005427	4885644	7161	D-003331-06	GAGACGAGGACACGTAATA	1362
TP73	NM_005427	4885644	7161	D-003331-07	CTGCAGAACCTGACCATTG	1363
TP73	NM_005427	4885644	7161	D-003331-08	GGCCATGCCTGTTTACAAG	1364
YWHAZ						
YWHAZ	NM_003406	21735623	7534	D-003332-05	GCAAGGAGCTGAATTATCC	1365
YWHAZ	NM_003406	21735623	7534	D-003332-06	TAAGAGATATCTGCAATGA	1366
YWHAZ	NM_003406	21735623	7534	D-003332-07	GACGGAAGGTGCTGAGAAA	1367
YWHAZ	NM_003406	21735623	7534	D-003332-08	AGAGCAAAGTCTTCTATTT	1368

Table IX

Gene Name	Accession #	GI#	Duplex #	Sequence	SEQ. ID NO.
AR	NM_000044	21322251	D-003400-01	GGAAGTCTGATCGTATCATT	1369
AR	NM_000044	21322251	D-003400-02	CAAGGGAGGTTACACCAAA	1370
AR	NM_000044	21322251	D-003400-03	TCAAGGAACTCGATCGTAT	1371
AR	NM_000044	21322251	D-003400-04	GAAATGATTGCACTATTGA	1372
ESR1	NM_000125	4503602	D-003401-01	GAATGTGCCTGGCTAGAGA	1373
ESR1	NM_000125	4503602	D-003401-02	CATGAGAGCTGCCAACCTT	1374
ESR1	NM_000125	4503602	D-003401-03	AGAGAAAGATTGGCCAGTA	1375

ESR1	NM_000125	4503602	D-003401-04	CAAGGAGACTCGCTACTGT	1376
ESR2	NM_001437	10835012	D-003402-01	GAACATCTGCTCAACATGA	1377
ESR2	NM_001437	10835012	D-003402-02	GCACGGCTCCATATACATA	1378
ESR2	NM_001437	10835012	D-003402-03	CAAGAAGATTCCCGGCTTT	1379
ESR2	NM_001437	10835012	D-003402-04	GGAAATGCGTAGAAGGAAT	1380
ESRRA	NM_004451	18860919	D-003403-01	GGCCTTCGCTGAGGACTTA	1381
ESRRA	NM_004451	18860919	D-003403-02	TGAATGCACTGGTGTCTCA	1382
ESRRA	NM_004451	18860919	D-003403-03	GCATTGAGCCTCTCTACAT	1383
ESRRA	NM_004451	18860919	D-003403-04	CCAGACAGCGGGCAAAGTG	1384
ESRRB	NM_004452	22035686	D-003404-01	TACCTGAGCTTACAAATTT	1385
ESRRB	NM_004452	22035686	D-003404-02	GCACTTCTATAGCGTCAAA	1386
ESRRB	NM_004452	22035686	D-003404-03	CAACTCCGATTCCATGTAC	1387
ESRRB	NM_004452	22035686	D-003404-04	GGACTCGCCACCCATGTTT	1388
ESRRG	NM_001438	4503604	D-003405-01	AAACAAAGATCGACACATT	1389
ESRRG	NM_001438	4503604	D-003405-02	TCAGGAAACTGTATGATGA	1390
ESRRG	NM_001438	4503604	D-003405-03	GAAGACCAGTCCAAATTAG	1391
ESRRG	NM_001438	4503604	D-003405-04	ATGAAGCGCTGCAGGATTA	1392
HNF4A	NM_000457	21361184	D-003406-01	CGACATCACTGGAGCATAT	1393
HNF4A	NM_000457	21361184	D-003406-02	GAAGGAAGCCGTCCAGAAT	1394
HNF4A	NM_000457	21361184	D-003406-03	CCAAGTACATCCCAGCTTT	1395
HNF4A	NM_000457	21361184	D-003406-04	GGACATGGCCGACTACAGT	1396
HNF4G	NM_004133	6631087	D-003407-01	GCACTGACATAAACGTTAA	1397
HNF4G	NM_004133	6631087	D-003407-02	ACAAAGAGATCCATGATGT	1398
HNF4G	NM_004133	6631087	D-003407-03	AGAGATCCATGATGTATAA	1399
HNF4G	NM_004133	6631087	D-003407-04	AAATGAACGTGACAGAATA	1400
HSAJ2425	NM_017532	8923776	D-003408-01	GAATGAATCTACACCTTG	1401
HSAJ2425	NM_017532	8923776	D-003408-02	GGAAATACGTGGAGACACT	1402
HSAJ2425	NM_017532	8923776	D-003408-03	CCAGATAACTACGGCGATA	1403
HSAJ2425	NM_017532	8923776	D-003408-04	TGGCGTACCTTCTCATTGA	1404
NR0B1	NM_000475	5016089	D-003409-01	CAGCATGGATGATATGATG	1405
NR0B1	NM_000475	5016089	D-003409-02	CTGCTGAGATTCATCAATG	1406
NR0B1	NM_000475	5016089	D-003409-03	ACAGATTCATCGAACTTAA	1407
NR0B1	NM_000475	5016089	D-003409-04	GAACGTGGCGCTCCTGTAC	1408
NR0B2	NM_021969	13259502	D-003410-01	GAATATGCCTGCCTGAAAG	1409
NR0B2	NM_021969	13259502	D-003410-02	GGAATATGCCTGCCTGAAA	1410
NR0B2	NM_021969	13259502	D-003410-03	CGTAGCCGCTGCCTATGTA	1411
NR0B2	NM_021969	13259502	D-003410-04	GCCATTCTCTACGCACTTC	1412
NR1D1	NM_021724	13430847	D-003411-01	CAACACAGGTGGCGTCATCTT	1413
NR1D1	NM_021724	13430847	D-003411-02	GGCATGGTGTTACTGTGTATT	1414
NR1D1	NM_021724	13430847	D-003411-03	CAACATGCATTCCGAGAAGTT	1415
NR1D1	NM_021724	13430847	D-003411-04	GCGCTTTGCTTCGTTGTTCTT	1416
NR1H2	NM_007121	11321629	D-003412-01	GAACAGATCCGGAAGAAGA	1417
NR1H2	NM_007121	11321629	D-003412-02	GAAGAACAGATCCGGAAGA	1418
NR1H2	NM_007121	11321629	D-003412-03	CTAAGCAAGTGCCTGGTTT	1419
NR1H2	NM_007121	11321629	D-003412-04	GCTAACAGCGGCTCAAGAA	1420

NR1H3	NM 005693	5031892	D-003413-01	GAACAGATCCGCCTGAAGA	1421
NR1H3	NM 005693	5031892	D-003413-02	GGAGATAGTTGACTTTGCT	1422
NR1H3	NM 005693	5031892	D-003413-03	GAGTTTGCCTTGCTCATTG	1423
NR1H3	NM 005693	5031892	D-003413-04	TGACTTTGCTAAACAGCTA	1424
NR1H4	NM 005123	4826979	D-003414-01	CAAGTGACCTCGACAACAA	1425
NR1H4	NM 005123	4826979	D-003414-02	GAAAGAATTTCGAAATAGTG	1426
NR1H4	NM 005123	4826979	D-003414-03	CAACAGACTCTTCTACATT	1427
NR1H4	NM 005123	4826979	D-003414-04	GAACCATACTCGCAATACA	1428
NR1I2	NM 003889	11863133	D-003415-01	GAACCATGCTGACTTTGTA	1429
NR1I2	NM 003889	11863133	D-003415-02	GATGGACGCTCAGATGAAA	1430
NR1I2	NM 003889	11863133	D-003415-03	CAACCTACATGTTCAAAGG	1431
NR1I2	NM 003889	11863133	D-003415-04	CAGGAGCAATTCGCCATTA	1432
NR1I3	NM 005122	4826660	D-003416-01	GGAAATCTGTACATCGTA	1433
NR1I3	NM 005122	4826660	D-003416-02	TCGCAGACATCAACACTTT	1434
NR1I3	NM 005122	4826660	D-003416-03	CCTCTTCGCTACACAATTG	1435
NR1I3	NM 005122	4826660	D-003416-04	GAACAGTTTGTGCAGTTTA	1436
NR2C1	NM 003297	4507672	D-003417-01	TGACAGCACTTGATCATAA	1437
NR2C1	NM 003297	4507672	D-003417-02	GGAAGGAAGTGACACCTA	1438
NR2C1	NM 003297	4507672	D-003417-03	GAGCACATCTTCAAACCTAC	1439
NR2C1	NM 003297	4507672	D-003417-04	GAAGAAATTGCACATCAAA	1440
NR2C2	NM 003298	4507674	D-003418-01	GAACAACGGTGACACTTCA	1441
NR2C2	NM 003298	4507674	D-003418-02	CTGATGAGCTCCAACATAA	1442
NR2C2	NM 003298	4507674	D-003418-03	CAACCTAAGTGAATCTTTG	1443
NR2C2	NM 003298	4507674	D-003418-04	GAAGACACCTACCGATTGG	1444
NR2E1	NM 003269	21361108	D-003419-01	GATCATATCTGAAATACAG	1445
NR2E1	NM 003269	21361108	D-003419-02	CAAGACTGCTTTCAGATAT	1446
NR2E1	NM 003269	21361108	D-003419-03	GTTAGATGCTACTGAATTT	1447
NR2E1	NM 003269	21361108	D-003419-04	CAATGTATCTCTATGAAGT	1448
NR2E3	NM 014249	7657394	D-003420-01	GAGAAGCTCCTTTGTGATA	1449
NR2E3	NM 014249	7657394	D-003420-02	GAAGCACTATGGCATCTAT	1450
NR2E3	NM 014249	7657394	D-003420-03	GAAGGATCCTGAGCACGTA	1451
NR2E3	NM 014249	7657394	D-003420-04	GAAGCTCCTTTGTGATATG	1452
NR2F1	NM 005654	20127484	D-003421-01	GAAACTCTCATCCGCGATA	1453
NR2F1	NM 005654	20127484	D-003421-02	TCTCATCCGCGATATGTTA	1454
NR2F1	NM 005654	20127484	D-003421-03	CAAGAAGTGCCTCAAAGTG	1455
NR2F1	NM 005654	20127484	D-003421-04	GGAACTTAACTTACACATG	1456
NR2F2	NM 021005	14149745	D-003422-01	GTACCTGTCCGGATATATT	1457
NR2F2	NM 021005	14149745	D-003422-02	CCAACCAGCCGACGAGATT	1458
NR2F2	NM 021005	14149745	D-003422-03	ACTCGTACCTGTCCGGATA	1459
NR2F2	NM 021005	14149745	D-003422-04	GGCCGTATATGGCAATTCA	1460
NR2F6	NM 005234	20070198	D-003423-01	CGACGCCTGTGGCCTCTCA	1461
NR2F6	NM 005234	20070198	D-003423-02	CAGCCGGTGTCCGAAGTGA	1462
NR2F6	NM 005234	20070198	D-003423-03	CAACCGTGACTGCCAGATC	1463
NR2F6	NM 005234	20070198	D-003423-04	GTAAGTCCGTCTCAAGAAG	1464

NR3C1	NM_000176	4504132	D-003424-01	GAGGACAGATGTACCACTA	1465
NR3C1	NM_000176	4504132	D-003424-02	GATAAGACCATGAGTATTG	1466
NR3C1	NM_000176	4504132	D-003424-03	GAAGACGATTCATTCCCTT	1467
NR3C1	NM_000176	4504132	D-003424-04	GGACAGATGTACCACTATG	1468
NR3C2	NM_000901	4505198	D-003425-01	GCAAACAGATGATCCAAGT	1469
NR3C2	NM_000901	4505198	D-003425-02	CAGCTAAGATTATCAGAA	1470
NR3C2	NM_000901	4505198	D-003425-03	GCACGAAAGTCAAAGAAGT	1471
NR3C2	NM_000901	4505198	D-003425-04	GGTATCCGGTCTTAGAATA	1472
NR4A1	NM_002135	21361341	D-003426-01	GAAGGAAGTTGTCCGAACA	1473
NR4A1	NM_002135	21361341	D-003426-02	CAGGAGAGTTTGACACCTT	1474
NR4A1	NM_002135	21361341	D-003426-03	CAGTGGCTCTGACTACTAT	1475
NR4A1	NM_002135	21361341	D-003426-04	GAAGGCCGCTGTGCTGTGT	1476
NR4A2	NM_006186	5453821	D-003427-01	GCAATGCGTTCGTGGCTTT	1477
NR4A2	NM_006186	5453821	D-003427-02	CGGCTACACAGGAGAGTTT	1478
NR4A2	NM_006186	5453821	D-003427-03	CCACGTGACTTTCAACAAT	1479
NR4A2	NM_006186	5453821	D-003427-04	GAATACAGCTCCGATTCT	1480
NR4A3	NM_006981	11276070	D-003428-01	CAAAGAAGATCAGACATTA	1481
NR4A3	NM_006981	11276070	D-003428-02	GATCAGACATTACTTATTG	1482
NR4A3	NM_006981	11276070	D-003428-03	CCAGAGATCTTGATTATTC	1483
NR4A3	NM_006981	11276070	D-003428-04	GAAGTTGTCCGTACAGATA	1484
NR5A1	NM_004959	20070192	D-003429-01	GATTTGAAGTTCCTGAATA	1485
NR5A1	NM_004959	20070192	D-003429-02	GGAGCGAGCTGCTGGTGTT	1486
NR5A1	NM_004959	20070192	D-003429-03	GGAGGTGGCCGACCAGATG	1487
NR5A1	NM_004959	20070192	D-003429-04	CAACGTGCCTGAGCTCATC	1488
NR5A2	NM_003822	20070161	D-003430-01	CCAAACATATGGCCACTTT	1489
NR5A2	NM_003822	20070161	D-003430-02	TCAGAGAAGTTAAGGTTGA	1490
NR5A2	NM_003822	20070161	D-003430-03	GGATCCATCTTCCTGGTTA	1491
NR5A2	NM_003822	20070161	D-003430-04	AAGAATACCTCTACTACAA	1492
NR6A1	NM_033334	15451847	D-003431-01	CAACGAACCTGTCTCATTT	1493
NR6A1	NM_033334	15451847	D-003431-02	GAAGAACTACACAGATTTA	1494
NR6A1	NM_033334	15451847	D-003431-03	GAAGATGGATACGCTGTGA	1495
NR6A1	NM_033334	15451847	D-003431-04	AAACGATACTGGTACATTT	1496
null	D16815	2116671	D-003432-01	GAAGAATGATCGAATAGAT	1497
null	D16815	2116671	D-003432-02	GAACATGGAGCAATATAAT	1498
null	D16815	2116671	D-003432-03	GAGGAGCTCTTGGCCTTTA	1499
null	D16815	2116671	D-003432-04	TAAACAACATGCACTCTGA	1500
PGR	NM_000926	4505766	D-003433-01	GAGATGAGGTCAAGCTACA	1501
PGR	NM_000926	4505766	D-003433-02	CAGCGTTTCTATCAACTTA	1502
PGR	NM_000926	4505766	D-003433-03	AGATAACTCTCATTCAGTA	1503
PGR	NM_000926	4505766	D-003433-04	GTAGTCAAGTGGTCTAAAT	1504
PPARA	NM_005036	7549810	D-003434-01	TCACGGAGCTCACGGAATT	1505
PPARA	NM_005036	7549810	D-003434-02	GAACATGACATAGAAGATT	1506
PPARA	NM_005036	7549810	D-003434-03	GGATAGTTCTGGAAGCTTT	1507
PPARA	NM_005036	7549810	D-003434-04	GACTCAAGCTGGTGTATGA	1508
PPARD	NM_006238	5453939	D-003435-01	GAGCGCAGCTGCAAGATTC	1509

PPARD	NM_006238	5453939	D-003435-02	GCATGAAGCTGGAGTACGA	1510
PPARD	NM_006238	5453939	D-003435-03	GGAAGCAGTTGGTGAATGG	1511
PPARD	NM_006238	5453939	D-003435-04	GCTGCAAGATTGAGAAGAA	1512
PPARG	NM_138712	20336234	D-003436-01	AGACTCAGCTCTACAATAA	1513
PPARG	NM_138712	20336234	D-003436-02	GATTGAAGCTTATCTATGA	1514
PPARG	NM_138712	20336234	D-003436-03	AAGTAACTCTCCTCAAATA	1515
PPARG	NM_138712	20336234	D-003436-04	GCATTTCTACTCCACATTA	1516
RARA	NM_000964	4506418	D-003437-01	GACAAGAAGCTGCATCATCA	1517
RARA	NM_000964	4506418	D-003437-02	GCAATACACTACGAACAA	1518
RARA	NM_000964	4506418	D-003437-03	GAACAACAGCTCAGAACAA	1519
RARA	NM_000964	4506418	D-003437-04	GAGCAGCAGTTCTGAAGAG	1520
RARB	NM_000965	14916493	D-003438-01	GCACACTGCTCAATCAATT	1521
RARB	NM_000965	14916493	D-003438-02	GCAGAAGTATTCAGAAGAA	1522
RARB	NM_000965	14916493	D-003438-03	GGAATGACAGGAACAAGAA	1523
RARB	NM_000965	14916493	D-003438-04	GCACAGTCCTAGCATCTCA	1524
RARG	NM_000966	21359851	D-003439-01	GAAATGACCGGAACAAGAA	1525
RARG	NM_000966	21359851	D-003439-02	TAGAAGAGCTCATCACCAA	1526
RARG	NM_000966	21359851	D-003439-03	CAAGGAAGCTGTGCGAAAT	1527
RARG	NM_000966	21359851	D-003439-04	TCAGTGAGCTGGCTACCAA	1528
RORA	NM_134261	19743902	D-003440-01	GGAAAGAGTTTATGTTCTA	1529
RORA	NM_134261	19743902	D-003440-02	CAAGATCTGTGGAGACAAA	1530
RORA	NM_134261	19743902	D-003440-03	GCACCTGACTGAAGATGAA	1531
RORA	NM_134261	19743902	D-003440-04	CCGAGAAGATGGAATACTA	1532
RORB	NM_006914	19743906	D-003441-01	GCACAGAACATCATTAAGT	1533
RORB	NM_006914	19743906	D-003441-02	CCACACCTATGAAGAAATT	1534
RORB	NM_006914	19743906	D-003441-03	GATCAAATTCTACTTCTGA	1535
RORB	NM_006914	19743906	D-003441-04	TCAAACAGATAAAGCAAGA	1536
RORC	NM_005060	19743908	D-003442-01	TAGAACAGCTGCAGTACAA	1537
RORC	NM_005060	19743908	D-003442-02	TCACCGAGGCCATTGAGTA	1538
RORC	NM_005060	19743908	D-003442-03	GAACAGCTGCAGTACAATC	1539
RORC	NM_005060	19743908	D-003442-04	CCTCATGCCACCTTGAATA	1540
RXRA	NM_002957	21536318	D-003443-01	TGACGGAGCTTGTGTCCAA	1541
RXRA	NM_002957	21536318	D-003443-02	CAACAAGGACTGCCTGATT	1542
RXRA	NM_002957	21536318	D-003443-03	GCAAGGACCTGACCTACAC	1543
RXRA	NM_002957	21536318	D-003443-04	GCAAGGACCGGAACGAGAA	1544
RXRB	NM_021976	21687229	D-003444-01	GCAAAGACCTTACATACTC	1545
RXRB	NM_021976	21687229	D-003444-02	GCAATCATTCTGTTTAATC	1546
RXRB	NM_021976	21687229	D-003444-03	TCACACCGATCCATTGATG	1547
RXRB	NM_021976	21687229	D-003444-04	GCAAACGGCTATGTGCAAT	1548
RXRG	NM_006917	21361386	D-003445-01	GGAAGGACCTCATCTACAC	1549
RXRG	NM_006917	21361386	D-003445-02	CCGGATCTCTGGTTAAACA	1550
RXRG	NM_006917	21361386	D-003445-03	GCGAGCCATTGTACTCTTT	1551
RXRG	NM_006917	21361386	D-003445-04	GAGCCATTGTACTCTTTAA	1552
THRA	NM_003250	20127451	D-003446-01	GGACAAAGACGAGCAGTGT	1553
THRA	NM_003250	20127451	D-003446-02	GGAAACAGAGGCGGAAATT	1554

THRA	NM_003250	20127451	D-003446-03	GTAAGCTGATTGAGCAGAA	1555
THRA	NM_003250	20127451	D-003446-04	GAACCTCCATCCCACCTAT	1556
THRB	NM_000461	10835122	D-003447-01	GAATGTCGCTTTAAGAAAT	1557
THRB	NM_000461	10835122	D-003447-02	GAACAGTCGTCGCCACATC	1558
THRB	NM_000461	10835122	D-003447-03	GGACAAGCACCAATAGTCA	1559
THRB	NM_000461	10835122	D-003447-04	GTGGAAAGGTTGACTTGGA	1560
VDR	NM_000376	4507882	D-003448-01	TGAAGAAGCTGAACTTGCA	1561
VDR	NM_000376	4507882	D-003448-02	GCAACCAAGACTACAAGTA	1562
VDR	NM_000376	4507882	D-003448-03	TCAATGCTATGACCTGTGA	1563
VDR	NM_000376	4507882	D-003448-04	CCATTGAGGTCATCATGTT	1564

Table X

Gene Symbol	Sense	SEQ. ID NO.
ABCB1	GACCAUAAAUGUAAGGUUU	1565
	UAGAAGAUCUGAUGUCAA	1566
	GAAUUGUUCACUUCAGUUA	1567
	GAAGAUCGCUACUGAAGCA	1568
ABCC1	GGAAGCAACUGCAGAGACA	1569
	GAUGACACCUCUCAACAAA	1570
	UAAAGUUGCUCAUCAAGUU	1571
	CAACGAGUCUGCCGAAGGA	1572
ABCG2	GCAGAUGCCUUCUUCGUUA	1573
	AGGCAAUUCUUCGUUAUUA	1574
	GGGAAGAAUUCUGGUCUAA	1575
	UGACUCAUCCCAACAUUUA	1576
KCNH2	CCGACGUGCUGCCUGAGUA	1577
	GAGAAGAGCAGCGACACUU	1578
	GAUCAUAGCACCUAAGAU	1579
	GCUAUUUACUGCUCUUAUU	1580
	UCACUGGGCUCCUUUAAUU	1581
	GUGCGAGCCUUCUGAAUUA	1582
	GCUAAGCUAUACUACUGUA	1583
	UGACGGCGCUCUACUUCAC	1584
KCNH1	GAGAUGAAUCCUUUGAAA	1585
	GAAGAACGCAUGAAACGAA	1586
	GAUAAAGACACGAUUGAAA	1587
	GCUGAGAGGUCUAUUUAAA	1588
CLCA1	GAACAACAAUGGCUAUGAA	1589
	GUACAUACCUGGCUGGAUU	1590

GAACAGCUCACAAGUAUUAU	1591
GGAAACGUGUGUCUAUAUU	1592

SLC6A1	Sense	
	GGAGGUGGGAGGACAGUUA	1593
	UCACAGCCCUGGUGGAUGA	1594
	GAAGCUGGCUCCUAUGUUC	1595
	GGUCAACACUACCAACAUG	1596

SLC6A2	Sense	
	GAACACAAGGUCAACAUUG	1597
	AGAAGGAGCUGGCCUAGUG	1598
	CGGAAACUCUUCACAUUUG	1599
	CAACAAAUUGACAACAAC	1600

SLC21A2	Sense	
	GUACAUCUCCAUCUUAUUU	1601
	GGAAGUGGCUGAGUUAUUA	1602
	GAAGGGAGGCUCAAUGUAA	1603
	GAAGGAAGUGGCUGAGUUA	1604

SLC21A3	Sense	
	GUAGAAACAGGAGCUAUUA	1605
	CAAGAUUACUGUCAACAA	1606
	GCACAAGAGUAUUUGGUAA	1607
	GCAAAUGUCCCUUCUGUAU	1608
	GCAUGACUCCUAUAUAUA	1609
	AAACAGCAAUUUCCCUUAA	1610
	GAAAUGCCUCUUCAGGAA	1611

SLC28A1	Sense	
	GUUCAUCGCUCUCCUCUUU	1612
	GGAUCAAGCUGUUUCUGAA	1613
	GGACUGCAGUUUGUACUUG	1614
	GAGUGAAACUGACCUAUGG	1615

SLC29A1	Sense	
	GAACGCUGCUCCCGUGGAA	1616
	GAAAGCCACUCUAUCAAG	1617
	GAAACCAGGUGCCUUCAGA	1618
	CCUCACAGCUGUAUUAUG	1619

SLC26A1	Sense	
	CCACGGAGCUGCUGGUCAU	1620
	GGGUUGACAUCUUAUUUGA	1621
	GCACGAGGGUCUCUGUGUU	1622
	GGCCAUCGCCUACUCAUUG	1623
	CAACACCCAUGGCAAUUA	1624
	GAGGAAAGAUCUUGCUGAU	1625
	GAGCAAGCGUCCUCCAAU	1626
	GCAACACCCAUGGCAAUUA	1627

SLC26A2	Sense	
	CCAAAGAACUCAUGAACA	1628
	ACAAGAACCUUCAGACUAA	1629
	GAAGGUAGAUAGAAGAAUG	1630

	GUUUUGAACUGUACUGUAA	1631
SLC4A4	Sense	
	GCAAUUCUCUUAUUUAUC	1632
	GGAAAGAUGUCCACUGAAA	1633
	GGACAAAGCCUUCUCAAU	1634
	GGAAUGGGAUCCAGCAAU	1635
GLRA1	Sense	
	UGAAAGCCAUUGACAUUUG	1636
	CAGACACGCUGGAGUUUAA	1637
	CAAUAGCGCUUUCUGGUUU	1638
	GCAGGUAGCAGAUGGACUA	1639
KLK1	Sense	
	UCAGAGUGCUGUCUUAUGU	1640
	CAACUUGUUUGACGACGAA	1641
	UGACAGAGCCUGCUGAUAC	1642
	AGGCGGCUCUGUACCAUUU	1643
ADAM2	Sense	
	GAAACAUGCUGUGAUUUUG	1644
	GCAGAUUUUCCUUAUAUA	1645
	CAACAGAGAUGCCAUGAUA	1646
	GAAAGGCGCUACAUUGAGA	1647
XPNPE P1	Sense	
	GACCUGAGCUUCCCAACAA	1648
	GCGACUGGCUCAACAAUUA	1649
	GAGAUUGCGUGGCUAUUUA	1650
	GACAGCAACUGGACACUUA	1651
GZMA	Sense	
	GGAAGAGACUCGUGCAAUG	1652
	GGAACCAUGUGCCAAGUUG	1653
	GAAGUAACUCCUCAUUCAA	1654
	GAACUCCUAUAGAUUUCUG	1655
CMKLR1	Sense	
	CAUAGAAGCUUUACCAAGA	1656
	GAAUGGAGGAUGAAGAUUA	1657
	GGUCA AUGCUCUAAGUGAA	1658
	GAGAGGACUUCUAUGAAUG	1659
CLN3	Sense	
	CAUCAUGCCUUCUGAAUAA	1660
	CAACAGCUCAUCACGAUUU	1661
	GCAACAACUUCUCUUAUGU	1662
	GGUCUUCGCUAGCAUCUCA	1663
CALCR	Sense	
	GGACCUAGCUGUUGUAAAG	1664
	GAAAGACCAUGCAUUUAAA	1665
	GCAGGAAGAUGUAUGCUUU	1666
	GAAUAAACCAGUAUCGUUA	1667

OXTR	Sense	
	GGACCCAGAUAUCCAAUA	1668
	GCAAUACUAUCCUACUGA	1669
	GAAUAUAGAUUAGCGUUUG	1670
	GAUGAGGCAUGACUACUAA	1671
EDG4	Sense	
	GCGAGUCUGUCCACUAUAC	1672
	GAGAACGGCCACCCACUGA	1673
	GAACGGCCACCCACUGAUG	1674
	GGUCAUUGCUGCUGUGUAC	1675
EDG5	Sense	
	UCCAGGAACACUAUAAUUA	1676
	GUGACCAUCUUCUCCAUA	1677
	CAUCCUCUGUUGCGCCAUAU	1678
	CCAACAAGGUCCAGGAACA	1679
EDG7	Sense	
	ACACUGAUACUGUCGAUGA	1680
	AAUAGGAGCAACACUGAUA	1681
	CAGCAGGAGUUACCUUGUU	1682
	GGACACCCAUGAAGCUAAU	1683
PTCH	Sense	
	GCACAGAACUCCACUCAA	1684
	GGACAGCAGUUCAUUGUUA	1685
	GAGAAGAGGCUAUGUUUAA	1686
	GGACAAACUUCGACCCUUU	1687
SMO	Sense	
	UCGCUACCCUGCUGUUAUU	1688
	GCUACAAGAACUACCGAUA	1689
	CAAGAAAGCUUCCUUAAC	1690
	GAGAAGAAUACAGUCAAU	1691
CASP3	Sense	
	CAAUUAUUCUGAAGAGCUA	1692
	GAACUGGACUGUGGCAUUG	1693
	GUGAGAAGAUGGUUAUUAUU	1694
	GAGGGUACUUUAAGACUAU	1695
CASP6	Sense	
	CAUGAGGUGUCAACUGUUA	1696
	GAAGUGAAAUGCUUUAUG	1697
	AAAU AUGGCUCUCCUUAAG	1698
	GCAAUCACAUUUAUGCAUA	1699
	CAACAUAAACUGAGGUGGAU	1700
	CAUGGUACAUAUCAAGAUUU	1701
CASP7	Sense	
	GAACUCUACUUCAGUCAAU	1704
	GGGCAAAUGCAUCAUAAUA	1703
	CAACAGAGGGAGUUUAAUA	1704
	GAACAAAGCCACUGACUGA	1705

CASP8

Sense	
GAAGUGAACUAUGAAGUAA	1706
CAACAAGGAUGACAAGAAA	1707
GGACAAAGUUUACCAAUG	1708
GAGGGUCGAUCAUCUAUUA	1709
GAAUAUAGAGGGCUUAUGA	1710
CAACGACUAUGAAGAAUUC	1711
GAAGUGAGCAGAUCAAGAAU	1712
GAGGAAAUCUCCAAUGCA	1713

CASP9

Sense	
CCAGGCAGCUGAUCAUAGA	1714
UCUCAGGUGUUGCCAAUA	1715
GAACAGCUGUAAUCUAUGA	1716
CCACUGGUCUGUAGGGAUU	1717

DVL1

Sense	
UCGUAAAGCUGUUGAUUUC	1718
GAGGAGAUUUUGAUGACA	1719
GUAAGCUGUUGAUUUCGA	1720
GAUCGUAAAGCUGUUGAUA	1721

DVL2

Sense	
AGACGAAGGUGAUUUACCA	1722
UGUGAGAGCUACCUAGUCA	1723
GAAGAAUUUCAGAUAGACA	1724
UAAUAGGCAUUUCCUCUUU	1725

PTEN

Sense	
GUGAAGAUCUUGACCAAUG	1726
GAUCAGCAUACACAAUUA	1727
GAAUGAACCUUCUGCAACA	1728
GGCGCUAUGUGUAUUUUA	1729

PDK1

Sense	
GUACAAAGCUGGUUAUUC	1730
GAAAGACUCCCAGUGUAUA	1731
GGAAGUCCAUCUCAUCGAA	1732
CAAAGACAUGACGACGUU	1733

PDK2

Sense	
GUAAGAGGAGACUGAAUG	1734
GGUCUGUGAUGGUCCUAA	1735
CAAAGAUGCCUACGACAUG	1736
GGGCGAUGCCUGAGGGUUA	1737

PPP2CA

Sense	
UCACACAAGUUUAUGGUUU	1738
CAACAGCCGUGACCACUUU	1739
UAACCAAGCUGCAAUCAUG	1740
GAACUUGACGAUACUCUAA	1741

CTNNA1

Sense	
GAAGAGAGGUCGUUCUAAG	1742
AAGCAGAUGUGCAUGAUUA	1743
UCUAAUAACUGCAGUGUUU	1744

	GUAAAGGGCCCUCUAAUAA	1745
CTNNA2	Sense	
	GAAAGAAUAUGCCCAAGUU	1746
	GAAGAAGAAUGCCACAAUG	1747
	GCAGGAAGAUUAUGAUGUG	1748
	AAAGAAAGCCCAUGUACUA	1749
HSPCA	Sense	
	GGGAAAGAGCUGCAUAUUA	1750
	GCUUAGAACUCUUUACUGA	1751
	UAUAAGAGCUUGACCAAUG	1752
	GCAGAUUUCUCUAUGAUUG	1753
DCTN2	Sense	
	CAACUCAUGUCCAAUACUG	1754
	GGAAUGAGCCAGAUGUUUA	1755
	GGAGACAGCUGUACGUUGU	1756
	UCCAAGAGCUGACAACUGA	1757
CD2	Sense	
	GUAAGGAGAAGCAAUAUAA	1758
	AAGAUGAGCUUCCAUGUA	1759
	GGACAUCUAUCUCAUCAUU	1760
	GACAAGAGCCACAGAGUA	1761
BAD	Sense	
	GUACUUCCCUCAGGCCUAU	1762
	GCUGUGCCUUGACUACGUA	1763
	GUACUUCCCUCAGGCCUAU	1764
	GGUCAGGUGCCUCGAGAUC	1765
SMAC	Sense	
	CAGCGUAACUUAUUCUUC	1766
	UAACUUAUUCUUCAGGUA	1767
	CAGCUGCUCUUAACCAUUU	1768
	GAUUGAAGCUAUUACUGAA	1769
	UAGAAGAGCUCCGUCAGAA	1770
	CCACAUUUGCGUUGAUUGA	1771
	GCGCAGGGCUCUCUACCUA	1772
MAP3K5	Sense	
	GAACAGCCUUCAAAUCAAA	1773
	GAUGUUCUCUACUAUGUUA	1774
	GCAAAUACUGGAAGGAUUA	1775
	CAGGAAAGCUCGUAAUUUA	1776
PVR	Sense	
	CCACACGGCUGACCUCAUA	1777
	CAGCAGAAUCCUCUUAUA	1778
	GCAGAAUCCUCUUAUAAA	1779
	GAUCGGGAUUUAUUUCUAU	1780
ERBB2	Sense	
	UGUGGGAGCUGAUGACUUU	1781
	UCACAGAGAUCUUGAAAGG	1782

UGGAAGAGAUCACAGGUUA	1783
GCUCAUCGCUCACAACCAA	1784

SOS1

Sense	
GAGCACCACUUCUAUGAUU	1785
CAAAGAAGCUGUUCAAUAU	1786
UGAAAGCCCUCUUUAUUA	1787
GAAAUAGCAUGGAGAAGGA	1788

BRCA1

Sense	
CCAUACAGCUUCAUAAUA	1789
GAAGAGAACUUAUCUAGUG	1790
GAAGUGGGCUCAGUAUUA	1791
GCAAGAUGCUGAUUCAUUA	1792
GAAGUGGGCUCAGUAUUA	1793
GAACGGACACUGAAUAUU	1794
GCAGAUAGUUCUACCAGUA	1795

CDKN1A

Sense	
GAACAAGGAGUCAGACAUU	1796
AAACUAGGCGGUUGAAUGA	1797
GAUGGAACUUCGACUUUGU	1798
GUAAACAGAUGGCACUUUG	1799

CDKN1B

Sense	
GGAAUGGACAUCUGUAUA	1800
GGAGAAAGAUGUCAACGU	1801
GAAUGGACAUCUGUAUAA	1802
GUAAACAGCUCGAAUUAAG	1803

SLC2A4

Sense	
CAGAUAGGCUCGGAAGAUG	1804
AGACUCAGCUCAGAAUAC	1805
GAUCGGUUCUUUCAUCUUC	1806
CAGGAUCGGUUCUUUCAUC	1807

NOS2A

Sense	
CCAGAUAAAGUGACAUAAAGU	1808
UAAGUGACCUGCUUUGUAA	1809
GAAGAGAGAUUCCAUGAA	1810
UGAAAGAGCUACAACAACAA	1811

FRAP1

Sense	
GAGCAUGCCGUCAAUAUA	1812
CAAGAGAACUCAUCAUAAG	1813
CCAAAGUGCUGCAGUACUA	1814
UAAGAAAGCUAUCCAGAUU	1815

FKBP1A

Sense	
GAAACAAGCCCUIUAAGUU	1816
GAAUUACUCUCCAAGUUGA	1817
CAGCACAAGUGGUAGGUUA	1818
GUUGAGGACUGAAUUACUC	1819
GAUGGCAGCUGUUUAAAUG	1820
GAGUAUCCUUCAGUGUUA	1821

TNFRSF
1A

Sense	
CAAAGGAACCUACUUGUAC	1822
GGAACCUACUUGUACAAUG	1823
GAACCUACUUGUACAAUGA	1824
GAGUGUGUCUCCUGUAGUA	1825

IL1R1

Sense	
GGACAAGAAUCAUUGGAUA	1826
GAACAAGCCUCCAGGAUUC	1827
GGACUUGUGUGCCCUUAUA	1828
GAACACAAAGGCACUAUAA	1829

IRAK1

Sense	
CGAAGAAAGUGAUGAAUUU	1830
GCUCUUUGCCCAUCUCUUU	1831
UGAAAGACCUGGUGGAAGA	1832
GCAAUUCAGUUUCUACAUC	1833

TRAF2

Sense	
GAAGACAGAGUUUUAAAC	1834
UCACGAAGACAGAGUUUU	1835
AGACAGAGUUUUAAACCA	1836
CACGAAGACAGAGUUUUUA	1837
GCUGAAGCCUGUCUGAUGU	1838

TRAF6

Sense	
CAAUGAUCUGAGGCAGUU	1839
GUUCAUAGUUUGAGCGUUA	1840
GGAGAAACCUGUUGUGAUU	1841
GGACAAAGUUGCUGAAAUC	1842
CAAUGAUCUGAGGCAGUU	1843
GGAGAAACCUGUUGUGAUU	1844
GGACAAAGUUGCUGAAAUC	1845
GUUCAUAGUUUGAGCGUUA	1846

TRADD

Sense	
UGAAGCACCUUGAUCUUUG	1847
GGGCAGCGCAUACCGUUU	1848
GAGGAGCGCUGUUUGAGUU	1849
GGACGAGGAGCGCUGUUUG	1850
GAGGAGCGCUGUUUGAGUU	1851
GGAUGUCUCUCUCCUUUU	1852
GCUCACUCCUUUCUACUAA	1853
UGAAGCACCUUGAUCUUUG	1854

FADD

Sense	
GCACAGAUUUUCCAUIIUC	1855
GCAGUCCUCUUAUUCUAA	1856
GAACUCAAGCUGCGUUUAU	1857
GGACGAAUUGAGAUAAUUA	1858

IKBKE

Sense	
UAAGAACACUGCUCAUGAA	1859
GAGGCAUCCUGAAGCAUUA	1860
GAAGGCGGCUGCAGAACUG	1861

GGAACAAGGAGAUCAUGUA	1862
---------------------	------

IKBKG

Sense	
CUAUCGAGGUCGUUAAAUU	1863
GAAUGCAGCUGGAAGAUUCU	1864
GCGGCGAGCUGGACUGUUU	1865
CCAGACCGAUGUGUAUUUA	1866

TNFRSF
5

Sense	
GGUCUCACCUCGCUAUGGU	1867
GAAAGCGAAUCCUAGACA	1868
GCACAAACAAGACUGAUGU	1869
GAAGGGCACCUCAGAAACA	1870
UCUCCCAACUUGUAUUAAA	1871

RELA

Sense	
UCAAGUGUCUCCAUCAUG	1872
UCAAGUGCCUAAUAGUAG	1873
GGAGUACCCUGAGGCUAUA	1874
GAUGAGAUCUCCUACUGU	1875

ARHA

Sense	
GAGCUGGGCUAAGUAAAUA	1876
GACCAAAGAUGGAGUGAGA	1877
GGAAGAAACUGGUGAUUGU	1878
GGCUGUAACUACUUUAUAA	1879

CDC42

Sense	
GGACAUUUGUUUGCCAUUU	1880
GGAGAACCAUAUACUCUUG	1881
GAACCA AUGCUUUCUCAUG	1882
GAAGACCUGUUAUGUAGAG	1883
GAUCAAGAAUUGCAAUAUC	1884
GAAAAGGGGUGACCUAGUA	1885
UGACAAACCUUAUGGAAA	1886

ROCK1

Sense	
GGAAUGAGCUUCAGAUGCA	1887
GGACACAGCUGUAAGAUUG	1888
GACAAGAGAUUACAGAUAA	1889
GAAGAAACA UCCCUAUUC	1890

PAK1

Sense	
GAGGGUGGUUUAUGAUUAA	1891
CAACAAAGAACAUCACUA	1892
GAAGAAAUUACACGGUUU	1893
UACAUGAGCUUUACAGUA	1894

PAK2

Sense	
GGUAGGAGAUGAAUUGUUU	1895
AGAAGGAACUGAUCAUUAA	1896
CUACAGACCUCCAUAUCA	1897
GAAACUGGCCAAACCGUUA	1898

PAK3

Sense	
-------	--

GAUUAUCGCUGCAAAGGAA	1899
GAGAGUGCCUGCAAGCUUU	1900
GACAAGAGGUGGCCAUAAA	1901
UAAAAUCGCUGUCUUGAGA	1902

PAK4	Sense	
	ACUAAGAGGUGAACAUUGUA	1903
	GAUCAUGAAUGUCCGAAGA	1904
	GAUGAGACCCUACUACUGA	1905
	CAGCAAAGGUGCCAAAGAU	1906

PAK6	Sense	
	UAAAGGCAGUUGUCCACUA	1907
	GAAGGGACCUGCUUUCUUG	1908
	GCAAAGACGUCCCUAAGAG	1909
	CCAAUGGGCUGGCUGCAA	1910

PAK7	Sense	
	GAGCACGGCUUUAUAAGU	1911
	CAAACUCCGUUAUGAUUA	1912
	GGAUAAAGUUGUCUGAUUU	1913
	GGAAUUGCCUCCAUAUA	1914

HDAC1	Sense	
	GGACAUCGCUGUGAAUUGG	1915
	AGAAAGAAGUCACCGAAGA	1916
	GGACAAGGCCACCCAAUGA	1917
	CCACAGCGAUGACUACAUU	1918

HDAC2	Sense	
	GCUGUAAAAUUAUGGCUUA	1919
	GCAAAGAAAGCUAGAAUUG	1920
	CAUCAGAGAGUCUUAUAUA	1921
	CCAAUGAGUUGCCAUUAUA	1922

CREBBP	Sense	
	GGCCAUAGCUAAUUAUUC	1923
	GCACAGCCGUUUACCAUGA	1924
	GGACAGCCCUUUAGUCAAG	1925
	GAACUGAUUCCUGAAAUAA	1926

BTRC	Sense	
	CACAUAAACUCGUAUCUUA	1927
	GAGAAGGCACUCAAGUUA	1928
	AGACAUAGUUUACAGAGAA	1929
	GCAGAGAGAUUUCUAACU	1930

RIPK2	Sense	
	GAACAUACCUGUAAAUCAU	1931
	GGACAUCGACCUGUUAUUA	1932
	UAAUUGAACUCCUACAUAG	1933
	GGAAUUAUCUCUGAACUA	1934

VAV1	Sense	
	GCAGAAAUACAUCUACUAA	1935
	GCUAUGAGCUGUUCUUCAA	1936

CGACAAAGCUCUACUCAUC	1937
GCUCAACCCUGGAGACAUU	1938

VAV2

Sense	
GGACAAGACUCGCAGAUUU	1939
GCUGAGCGCUUUGCAAUAA	1940
CAAGAAGUCUCACGGGAAA	1941
UCACAGAGGCCAAGAAAUU	1942

GRB2

Sense	
UGGAAGCCAUCGCCAAUA	1942
CAUCAGUGCAUGACGUUUA	1943
UGAAUGAGCUGGUGGAUUA	1944
UGCCAAAACUUACCUAUA	1945

PLCG1

Sense	
GAGCUGCACUCCAAUGAGA	1946
GAAACCAAGCCAUUAAUGA	1947
CCAAGGAGCUACUGACAUU	1948
AGAGAAACAUGGCCCAUA	1949

ITGB1

Sense	
CCACAGACAUUUACAUAUA	1950
GAAGGGAGUUUGCUAAAUU	1951
GAACAGAUCUGAUGAAUGA	1952
CAAGAGAGCUGAAGACUAU	1953

ITGA4

Sense	
GCAUAUAUAUUCAGCAUUG	1954
CAACUUGACUGCAGUAUUG	1955
GAACUUAACUUCCAUGUU	1956
GACAAGACCUGUAGUAAUU	1957

STAT1

Sense	
AGAAAGAGCUUGACAGUAA	1958
GGAAGUAGUUCACAAAUA	1959
UGAAGUAUCUGUAUCCAAA	1960
GAGCUUCACUCCCUUAGUU	1961

KRAS2

Sense	
UAAGGACUCUGAAGAUGUA	1962
GACAAAGUGUGUAAUUAUG	1963
GCUCAGGACUUAGCAAGAA	1964
GAAACUGAAUACCUAAGAU	1965
GAAACUGAAUACCUAAGAU	1966
UAAGGACUCUGAAGAUGUA	1967
GACAAAGUGUGUAAUUAUG	1968
GCUCAGGACUUAGCAAGAA	1969

HRAS

Sense	
CCAUCCAGCUGAUCCAGAA	1970
GAACCCUCCUGAUGAGAGU	1971
GAGGACAUCCACCAGUACA	1972

BRAF

Sense	
GAUUAGAGACCAAGGAUUU	

CCACUGAUGUGUGUAAUU	1973
CAAUAGAACCUGUCAAUAU	1974
GAAGACAGGAAUCGAAUGA	1975

ELK1

Sense	
GAUGUGAGUAGAAGAGUUA	1975
GGAAGAAUUUGUACCAUUU	1976
GAACGACCUUUCUUUCUUU	1977
GGAGUCAUCUCUCCUAUA	1978

RALGDS

Sense	
GGAGAAGCCUCACCUCUUG	1979
GCAGAAAGGACUCAAGAUU	1980
GAGAACAACUACUCAUUGA	1981
GAACUUCUCGUCACUGUAU	1982

PRKCA

Sense	
GGAUUGUUCUUUCUUAUA	1983
GAAGGGUUCUCGUAUGUCA	1984
GAAGAAGGAUGUGGUGAUU	1985
GGACUGGGAUCGAACAACA	1986

MAP2K4

Sense	
GGACAGAAGUGGAAAUUU	1987
UCAAGAGGGUGAACAUUAA	1988
GACCAAUCUCAGUUGUUU	1989
GGAGAAUGGUGCUGUUUAA	1990

MAP2K7

Sense	
GAAGAGACCAAAGUAUAAU	1991
GAAGACCGGCCACGUCAUU	1992
GGAAGAGACCAAAGUAUAA	1993
GCAUUGAGAUUGACCAGAA	1994
UGAGAGAACGAGAAAGUUG	1995
GUGAAACCCUGUCUGCAUU	1996
GGAUCUCUCUCAACAACUA	1997
ACAACUAGGUGAACACAUA	1998

MAPK8

Sense	
UCACAGUCCUGAAACGAUA	1999
GAUUGGAGAUUCUACAUUC	2000
GCUCAUGGAUGCAAUCUU	2001
GAAGCUAAGCCGACCAUUU	2002

MAPK9

Sense	
AAAGAGAGCUUAUCGUGAA	2003
GAUGAUAGGUUAGAAAUAG	2004
ACAAAGAAGUCAUGGAUUG	2005
GGAGCUGGAUCAUGAAAGA	2006

AIF1

Sense	
GAAAAGGGAUGAUGGGAUU	2007
CCUAGACGAUCCCAAUAU	2008
GAGCCAAACCAGGGAUUUA	2009
UGAAACGAAUGCUGGAGAA	2010
UCACUCACCCAGAGAAUA	2011

CCAAGAAAGCUAUCUCUGA	2012
AGACUCACCUAGAGCUAAA	2013

BBC3

Sense	
CCUGGAGGGUCCUGUACAA	2014
GAGCAAUGAGCCAAACGU	2015
GGAGGGUCCUGUACAAUCU	2016
GACUUUCUCUGCACCAUGU	2017

BCL2L1

Sense	
CCAGGGAGCUUGAAAGUUU	2018
AAAGUGCAGUUCAGUAAUA	2019
GAGAAUCACUAACCAGAGA	2020
GAGCCCAUCCCUAUUAUAA	2021

BCL2L11

Sense	
GAGACGAGUUUAACGCUUA	2022
AAAGCAACCUUCUGAUGUA	2023
CCGAGAAGGUAGACAAUUG	2024
GCAAAGCAACCUUCUGAUG	2025
AGACAGAGCCACAAGGUAA	2026
GCAAGGAGGUUAGAGAAAU	2027
CAAGGAGGUUAGAGAAUA	2028
UCUUACGACUGUUACGUUA	2029

BID

Sense	
GAAGACAUCAUCCGGAAUA	2030
CAACAGCGUCCUAGAGAA	2031
GAAAUGGGAUGGACUGAAC	2032
ACGAUGAGCUGCAGACUGA	2033

BIRC2

Sense	
GAAAGAAGCCUGCAUAUAA	2034
GAAAUUGACUCUACAUUGU	2035
ACAAAUAGCACUUAGGUUA	2036
GAAUACACCUGUGGUUAAA	2037

BIRC3

Sense	
GGAGAUGCCUGCCAUUAAA	2038
UCAAUGAUCUUGUGUUAGA	2039
GAAAGAACAUGUAAAGUGU	2040
GAAGAAAGAACAUGUAAAG	2041

BIRC4

Sense	
GUAGAUAGAUGGCAUAUUG	2042
GAGGAGGGCUAACUGAUUG	2043
GAGGAACCCUGCCAUGUAU	2044
GCACGGAUCUUUACUUUUG	2045

BIRC5

Sense	
GGCGUAAGAUGAUGGAUUU	2046
GCAAAGGAAACCAACAAUA	2047
GCACAAAGCCAUUCUAAGU	2048
CAAAGGAAACCAACAAUAA	2049

BRCA1

Sense	
-------	--

CCAUACAGCUUCAUAAUA	2050
GAAGAGAACUUAUCUAGUG	2051
GAAGUGGGCUCCAGUAUUA	2052
GCAAGAUGCUGAUUCAUUA	2053
CCAUACAGCUUCAUAAUA	2054

CARD4	Sense	
	GAAAGUUAUGUCAAGGAA	2055
	GAGCAACACUGGCAUAACA	2056
	UAACAGAGAUUUGCCUAAA	2057
	GCGAAGAGCUGACCAAUA	2058

CASP10	Sense	
	CAAAGGGUUCUCUGUUUA	2059
	GAAAUGACCUCUCCUAAGUU	2060
	GAAGGCAGCUGGUUAUUC	2061
	GACAUGAUCUCCUUCUGA	2062
	GCACUCUUCUGUCCCUUA	2063

CASP2	Sense	
	GUAUUAAACUCUCCUUUGA	2064
	GCAAGGAGAUGUCUGAAUA	2065
	CAACUCCCUGAUCUUUAA	2066
	GCUCAAGAUGUAAUGUAG	2067

CDKN1A	Sense	
	GAACAAGGAGUCAGACAUU	2068
	AAACUAGGCGGUUGAAUGA	2069
	GAUGGAACUUCGACUUUGU	2070
	GUAAACAGAUGGCACUUUG	2071

CFLAR	Sense	
	GAUGUGUCCUCAUUAAUUU	2072
	GAAGAGAGAUACAAGAUGA	2073
	GAGCAUACCUGAAGAGAGA	2074
	GCUAUGAAGUCCAGAAAUU	2075

CLK2	Sense	
	GUGAAUAUGUGAAAUGUG	2076
	AAAGCAUGCUAGAGUAUGA	2077
	UUAAGAAUGUGGAGAAGUA	2078
	GAUAACAAGCUGACACAU	2079

CLSPN	Sense	
	GGACGUAAUUGAUGAAGUA	2080
	GCAGAUGGGUUCUAAAUG	2081
	CAAUGAGGUUGAGGAAU	2082
	GGAAAUACCUGGAGGAUGA	2083

CSNK2A 1	Sense	
	GAUCCACGUUUCUAUGUA	2084
	GCAUUUAGGUGGAGACUUC	2085
	GAUGUACGAUUUAGUUUG	2086
	UGAAUUAGAUCACGUUUC	2087

CTNNB1	Sense	
	GCACAAGAAUGGAUCACAA	2088
	GCUGAAACAUGCAGUUGUA	2089
	GUACGUACCAUGCAGAAUA	2090
	GAACUUGCAUUGUGAUUGG	2091

CXCR4	Sense	
	GAAGCAUGACGGACAAGUA	2092
	GAACAUUCCAGAGCGUGUA	2093
	GUUCUUAGUUGCUGUAUGU	2094
	CAUCAUGGUUGGCCUUAUC	2095

CXCR6	Sense	
	GGAACAAACUGGCAAAGCA	2096
	GAUCAGAGCAGCAGUGAAA	2097
	GGGCAAAACUGAAUUAUAA	2098
	GAUCUCAGGUUCUCCUUGA	2099

DAXX	Sense	
	CUACAGAUCCCAAUGAAA	2100
	GCUACAAGCUGGAGAAUGA	2101
	GGAAACAGCUAUGUGGAAA	2102
	GGAGUUGGAUCUCUCAGAA	2103

GAS41	Sense	
	GUAGUAAGCUAAACUGAAA	2104
	GACAAUAUGUUAAGAGAA	2105
	GACAACAUCUCGUCAGCUA	2106
	UAUAUGAUGUGUCCAGUAA	2107

GTSE1	Sense	
	CAAAGAAGCUCACUACUG	2108
	GAACAGCCCUAAGUGGUU	2109
	GAACAUGGAUGACCCUAAG	2110
	GGGCAAAGCUAAAUCAAGU	2111

HDAC3	Sense	
	GGAAAGCGAUGUGGAGAUU	2112
	CCAAGACCGUGGCCUAUUU	2113
	AAAGCGAUGUGGAGAUUUA	2114
	GUGAGGAGCUUCCCUAUG	2115

HDAC5	Sense	
	GAAUCCUCUUGUCGAAGU	2116
	GUUAUUAGCACCUUUAAGA	2117
	GGAGGGAGGCCAUGACUUG	2118
	CAGGAGAGCUCAAGAAUGG	2119
	GGAUAUGGAUUUCAGUUA	2120
	GGAAGUCGGUGCCUUGGUU	2121
	GGAAGGAGAGGACUGGUUU	2122

HEC	Sense	
	GCAGAUACUUGCACGGUUU	2123
	GAGUAGAACUAGAAUGUGA	2124
	GCGAAUAAAUCAUGAAAGA	2125
	GAAGAUGGAUUUAUGCAUA	2126

HIST1H2	Sense	
AA	GGCAAUGCGUCUCGCGAUA	2127
	GAUCCGCAAUGAUGAGGAA	2128
	GCAAUGCGUCUCGCGAUA	2129
	GAGGAACUCAUAAGCUUU	2130

LMNB1	Sense	
	AAUAGAAGCUGUGCAAUUA	2131
	CAACUGACCUCUAUCUGGAA	2132
	GAAGGAAUCUGAUCUUAU	2133
	GGGAAGGGUUCUCUAUUA	2134

LMNB2	Sense	
	GGAGGUUCAUUGAGAAUUG	2134
	GGCAAUAGCUCACCGUUUA	2135
	CAAAUACGCUUAGCUGUGU	2136
	GGAGAUCGCCUACAAGUUC	2137

MYB	Sense	
	GCAGAAACACUCCAUAUUA	2138
	GUAAAUACGUGAAUGCAUU	2139
	GCACUGAACUUUUGAGAU	2140
	GAAGAACAGUCAUUUGAUG	2141

MYT1	Sense	
	GAGGUGAGCUGUUAAUUA	2142
	GCAGGGUGAUUCCUAUA	2143
	GGGAGAAGAUUUUAAUUG	2144
	CAACUUCUCUCCUGAACUU	2145

NFKBIB	Sense	
	GGACACGGCACUGCACUUG	2146
	GCACUUGGCUGUGAUUCAU	2148
	GAGACGAGGGCGAUGAAUA	2149
	CAUGAACCCUCCUGGAUU	2150

NFKBIA	Sense	
	GAACAUGGACUUGUAUAUU	2151
	GAUGUGGGGUGAAAAGUUA	2152
	GGACGAGAAAGAUCAUUGA	2153
	AGGACGAGCUGCCCUAUGA	2154

NFKBIE	Sense	
	GAAGGGAAGUUUCAGUAAC	2155
	GGAAGGGAAGUUUCAGUAA	2156
	GGAAACUGCUGCUGUGUAC	2157
	GAACCAACCACUCAUGGAA	2158

NUMA1	Sense	
	GGGAACAGUUUGAAUAUAA	2159
	GCAGUAGCCUGAAGCAGAA	2160
	CGAGAAGGAUGCACAGAU	2161
	GCAAGAGGCUGAGAGGAAA	2162

NUP153	Sense	
--------	-------	--

GAAGACAAAUGAAAGCUAA	2163
GAUAAAGACUGCUGUUAGA	2164
GAGGAGAGCUCUAAUAUUA	2165
GAGGAAGCCUGAUUAAAGA	2166

OPA1	Sense	
	GAAAGAGCAUGAUGACAUAA	2167
	GAGGAGAGCUCUAUUUAUGU	2168
	GAAACUGAAUGGAAGAAUA	2169
	AAAGAAGGCUGUACCGUUA	2170

PARVA	Sense	
	CUACAUGUCUUUGCUCUUA	2171
	GCUAAGUCCUGUAAGAAUA	2172
	CAAAGGCAAUGUACUGUUU	2173
	GAACAAUGGUGGAUCCAAA	2174

PIK3CG	Sense	
	AAGUUCAGCUUCUCUAUUA	2175
	GAAGAAUUCUCUGAUGGAU	2176
	GAACACCUUUACUCUAUAA	2177
	GCAUGGAGCUGGAGAACUA	2178

PRKDC	Sense	
	GAUGAAAGCUCUAAAGAUG	2179
	GAAAGGAGGUUCUAAACUA	2180
	GGAAGAAGCUCAUUUGAUU	2181
	GCAAAGAGGUGGCAGUUA	2182

RASA1	Sense	
	GGAAGAAGAUCCACAUGAA	2183
	GAACAUACUUUCAGAGCUU	2184
	GAACAAUCUUUGCUGUAUA	2185
	UACAGAACUGCUUCAACA	2186

SLC9A1	Sense	
	GAAGAGAUCCACACACAGU	2187
	UCAAUGAGCUGCUGCACA	2188
	GAAGAUAGGUUCCAUGUG	2189
	GAAUUACCCUCCUCAUCU	2190

TEGT	Sense	
	CUACAGAGCUUCAGUGUGA	2191
	GAACAUUUUGAUCGAAAG	2192
	GAGCAAACCUAGAUAGGA	2193
	GCAUUGAUCUCUUCUAGA	2194

TERT	Sense	
	GGAAGACAGUGGUGAACUU	2195
	GCAAAGCAUUGGAAUCAGA	2196
	GAGCUGACGUGGAAGAUGA	2197
	GAACGGGCCUGGAACCAUA	2198

TNFRSF 6	Sense	
	GAUACUAAACUGCUCUCAGA	2199

GAAAGAAUGGUGUCAAUGA	2200
UCAAUAAUGUCCCAUGUAA	2201
UCAUGAAUCUCCAACCUUA	2202
GAUGUUGACUUGAGUAAAU	2203

TOP1	Sense	
	GAAAGGAAAUGACUAAUGA	2204
	GAAGAAGGCUGUUCAGAGA	2205
	GGAAGUAGCUACGUUCUUU	2206
	GGACAUAAAGUGGAAAGAAG	2207

TOP2A	Sense	
	GAAAGAGUCCAUCAGAUUU	2208
	CAAACUACAUUGGCAUUUA	2209
	AAACAGACAUGGAUGGAUA	2210
	CGAAAGGAAUGGUUAACUA	2211

TOP3A	Sense	
	CCAGAAAUCUCCACAGAA	2212
	GAAACUAUCUGGAUGUGUA	2213
	CCACAAAGAUGGUAUCGUA	2214
	GGAAAUGGCUGUGGUAACA	2215

TOP3B	Sense	
	GAGACAAGAUGAAGACUGU	2216
	GCACAUGGGCUGCGUCUUU	2217
	CCAGUGCGCUUCAAGAUGA	2218
	GAACAUCUGCUUUGAGGUU	2219

WEE1	Sense	
	GGUAUUGCCUUGUGAAUUU	2220
	GCAGAACAAUACGAAUAG	2221
	GUACAUAGCUGUUUGAAAU	2222
	GCUGUAAACUUGUAGCAUU	2223

- In addition, to identifying functional siRNA against gene families or pathways, it is possible to design duplexes against genes known to be involved in specific diseases. For example when dealing with human disorders associated with allergies, it will be beneficial to develop siRNA against a number of genes including but not limited to:
- the interleukin 4 receptor gene
(SEQ. ID NO. 2224: UAGAGGUGCUCAUUCAUUU,
- 10 SEQ. ID NO. 2225: GGUAUAAGCCUUUCCAAGA,
SEQ. ID NO. 2225: ACACACAGCUGGAAGAAAU,
SEQ. ID NO. 2226: UAACAGAGCUUCCUUAGGU),

the Beta-arrestin-2

(SEQ. ID NO. 2227: GGAUGAAGGAUGACGACUA,
 SEQ. ID NO. 2228: ACACCAACCUCUAUGAAUU,
 SEQ. ID NO. 2229: CGAACAAGAUGACCAGGUA,
 SEQ. ID NO. 2230: GAUGAAGGAUGACGACUAU,),

5

the interferon-gamma receptor 1 gene

(SEQ. ID NO. 2231: CAGCAUGGCUCUCCUCUUU,
 SEQ. ID NO. 2232: GUAAAGAACUAUGGUGUUA,
 SEQ. ID NO. 2233: GAAACUACCUGUUACAUA,
 10 SEQ. ID NO. 2234: GAAGUGAGAUCAGUAUAA),

the matrix metalloproteinase MMP-9

(SEQ. ID NO. 2235: GGAACCAGCUGUAUUUGUU,
 SEQ. ID NO. 2236: GUUGGAGUGUUUCUAAUAA,
 15 SEQ. ID NO. 2237: GCGCUGGGCUUAGAUCUU,
 SEQ. ID NO. 2238: GGAGCCAGUUUGCCGGAUA),

the Slc11a1 (Nramp1) gene

(SEQ. ID NO. 2239: CCAAUGGCCUGCUGAACAA,
 20 SEQ. ID NO. 2240: GGGCCUGGCUUCCUCAUGA,
 SEQ. ID NO. 2241: GGGCAGAGCUCCACCAUGA,
 SEQ. ID NO. 2242: GCACGGCCAUUGCAUUCAA),

SPINK5

25 (SEQ. ID NO. 2243: CCAACUGCCUGUUCUAAUAA,
 SEQ. ID NO. 2244: GGAUACAUGUGAUGAGUUU,
 SEQ. ID NO. 2245: GGACGAAUGUGCUGAGUAU,
 SEQ. ID NO. 2246: GAGCUUGUCUUAUUUGCUA,),

30 the CYP1A2 gene

(SEQ. ID NO. 2247: GAAAUGCUGUGUCUUCGUA,
 SEQ. ID NO. 2248: GGACAGCACUUCCCUGAGA,
 SEQ. ID NO. 2249: GAAGACACCACCAUUCUGA,
 SEQ. ID NO. 2250: GGCCAGAGCUUGACCUUCA),

thymosin-beta4Y

(SEQ. ID NO. 2251: GGACAGGCCUGCGUUGUUU,
 SEQ. ID NO. 2252: GGAAAGAGGAAGCUCAUGA,
 5 SEQ. ID NO. 2253: GCAAACACGUUGGAUGAGU,
 SEQ. ID NO. 2254: GGACUAUGCUGCCCUUUUG,

activin A receptor IB

(SEQ. ID NO. 2255: ACAAGACGCUCCAGGAUCU,
 10 SEQ. ID NO. 2254: GCAACAGGAUCGACUUGAG,
 SEQ. ID NO. 2255: GAAGCUGCGUCCCAACAUC,
 SEQ. ID NO. 2256: GCAUAGGCCUGUAAUCGUA,
 SEQ. ID NO. 2257: UCAGAGAGUUCGAGACAAA,
 SEQ. ID NO. 2258: UGCGAAAGGUUGUAUGUGA,
 15 SEQ. ID NO. 2259: GCAACAGGAUCGACUUGAG,
 SEQ. ID NO. 2260: GAAUAGCGUUGUGUGUUUAU,
 SEQ. ID NO. 2261: UGAAUAGCGUUGUGUGUUA,
 SEQ. ID NO. 2262: GGGAUCAGUUUGUUGAAUA,
 SEQ. ID NO. 2263: GAGCCUGAAUCAUCGUUUA,),

20

- 2'

ADAM33

(SEQ. ID NO. 2264: GGAAGUACCUGGAACUGUA,
 SEQ. ID NO. 2265: GGACAGAGGGAACCAUUUA,
 SEQ. ID NO. 2266: GGUGAGAGGUAGCUCCUAA,
 25 SEQ. ID NO. 2267: AAAGACAGGUGGCCACUGA),

the TAP1 gene

(SEQ. ID NO. 2268: GAAAGAUGAUCAGCUAUUU,
 SEQ. ID NO. 2269: CAACAGAACCAGACAGGUA,
 30 SEQ. ID NO. 2270: UGAGAAAUGUUCAGAAUGU,
 SEQ. ID NO. 2271: UACCUUCACUCGAAACUUA,

COX-2

(SEQ. ID NO. 2272: GAACGAAAGUAAAGAUGUU,

SEQ. ID NO. 2273: GGACUUAUGGGUAAUGUUA,
 SEQ. ID NO. 2274: UGAAAGGACUUAUGGGUAA,
 SEQ. ID NO. 2275: GAUCAGAGUUCACUUCUU),

5 ADPRT

(SEQ. ID NO. 2276: GGAAAGAUGUUAAGCAUUU,
 SEQ. ID NO. 2277: CAUGGGAGCUCUUGAAUA,
 SEQ. ID NO. 2278: GAACAAGGAUGAAGUGAAG,
 SEQ. ID NO. 2279: UGAAGAAGCUCACAGUAAA,),

10

HDC

(SEQ. ID NO. 2280: CAGCAGACCUUCAGUGUGA,
 SEQ. ID NO. 2281: GGAGAGAGAUGGUGGAUUA,
 SEQ. ID NO. 2282: GUACAGAGCUGGAGAUGAA,
 15 SEQ. ID NO. 2283: GAACGUCCCUUCAGUCUGU),

HnmT

(SEQ. ID NO. 2284: CAAAUUCUCUCCAAAGUUC,
 SEQ. ID NO. 2285: GGAUAUAUCUGACUGCUUU,
 20 SEQ. ID NO. 2286: GAGCAGAGCUUGGGAAAGA,
 SEQ. ID NO. 2287: GAUAUGAGAUGUAGCAAAU),

GATA-3

(SEQ. ID NO. 2288: GAACUGCUUUCUUCGUUU,
 25 SEQ. ID NO. 2289: GCAGUAUCAUGAAGCCUAA,
 SEQ. ID NO. 2290: GAAACUAGGUCUGAUUUC,
 SEQ. ID NO. 2291: GUACAGCUCCGGACUCUUC),

Gab2

30 (SEQ. ID NO. 2292: GCACAACCAUUCUGAAGUU,
 SEQ. ID NO. 2293: GGACUUAUGAUGCCCAGAUG,
 SEQ. ID NO. 2294: GAAGGUGGAUUCUAGGAAA,
 SEQ. ID NO. 2295: GGACUAGCCCUGCUGUUUA), and

STAT6

(SEQ. ID NO. 2296: GAUAGAAACUCCUGCUAAU,
 SEQ. ID NO. 2297: GGACAUUUUAUUCCCAGCUA,
 SEQ. ID NO. 2298: GGACAGAGCUACAGACCUA,
 5 SEQ. ID NO. 2299: GGAUGGCUCUCCACAGAU).

In addition, rationally designed siRNA or siRNA pools can be directed against genes involved in anemia, hemophila or hypercholesterolemia. Such genes would include, but are not be limited to:

10 APOA5

(SEQ. ID NO. 2300: GAAAGACAGCCUUGAGCAA,
 SEQ. ID NO. 2301: GGACAGGGAGGCCACCAAA,
 SEQ. ID NO. 2302: GGACGAGGCUUGGGCUUUG,
 SEQ. ID NO. 2303: AGCAAGACCUCAACAAUAU),

15

HMG-CoA reductase

(SEQ. ID NO. 2304: GAAUGAAGCUUUGCCCUUU,
 SEQ. ID NO. 2305: GAACACAGUUUAGUGCUUU,
 SEQ. ID NO. 2306: UAUCAGAGCUCUAAUGUU,
 20 SEQ. ID NO. 2307: UGAAGAAUGUCUACAGAU),

NOS3

(SEQ. ID NO. 2308: UGAAGCACCUGGAGAAUGA,
 SEQ. ID NO. 2309: CGGAACAGCACAAGAGUUA,
 25 SEQ. ID NO. 2310: GGAAGAAGACCUUAAAGA,
 SEQ. ID NO. 2309: GCACAAGAGUUUAAGAUC),

ARH

(SEQ. ID NO. 2310: CGAUACAGCUUGGCACUUU,
 30 SEQ. ID NO. 2311: GAGAAGCGCUGCCCUGUGA,
 SEQ. ID NO. 2312: GAAUCAUGCUGUUCUCUUU,
 SEQ. ID NO. 2313: GGAGUAACCGGACACCUUA),

CYP7A1

(SEQ. ID NO. 2314: UAAGGUGACUCGAGUGUUU,
 SEQ. ID NO. 2315: AAACGACACUUUCAUCAA,
 SEQ. ID NO. 2316: GGACUCAAGUAAAAGUAUU,
 SEQ. ID NO. 2317: GUAAUGGACUCAAGUAAA),

5

FANCA

(SEQ. ID NO. 2318: GGACAUCACUGCCCACUUC,
 SEQ. ID NO. 2319: AGAGGAAGAUGUUCACUUA,
 SEQ. ID NO. 2320: GAUCGUGGCUCUUCAGGAA,
 10 SEQ. ID NO. 2321: GGACAGAGGCAGAUAAAGAA),

FANCG

(SEQ. ID NO. 2322: GCACUAAGCAGCCUUCAUG,
 SEQ. ID NO. 2323: GCAAGCAGGUGCCUACAGA,
 15 SEQ. ID NO. 2324: GGAAUUAGAUGCUCCAUG,
 SEQ. ID NO. 2325: GGACAUCUCUGCCAAAGUC),

ALAS

(SEQ. ID NO. 2326: CAAUAUGCCUGGAAACUAU,
 20 SEQ. ID NO. 2327: GGUUAAGACUCACCAGUUC,
 SEQ. ID NO. 2328: CAACAGGACUUUAGGUUCA,
 SEQ. ID NO. 2329: GCAUAAGAUUGACAUCAUC),

PIGA

25 (SEQ. ID NO. 2330: GAAAGAGGGCAUAAGGUUA,
 SEQ. ID NO. 2331: GGACUGAUCUUUAAACUAU,
 SEQ. ID NO. 2332: UCAAAUGGCUUACUUCAUC,
 SEQ. ID NO. 2333: UCUAAGAACUGAUGUCUAA), and

30 factor VIII

(SEQ. ID NO. 2334: GCAAAUAGAUCUCCAUIUAC,
 SEQ. ID NO. 2335: CCAGAUUAUGUCGUUCUUUA,
 SEQ. ID NO. 2336: GAAAGGCUGUGCUCUCAA,
 SEQ. ID NO. 2337: GGAGAAACCUGCAUGAAAG,

SEQ. ID NO. 2338: CUUGAAGCCUCCUGAAUUA,
 SEQ. ID NO. 2339: GAGGAAGCAUCCAAAGAUU,
 SEQ. ID NO. 2340: GAUAGGAGAUACAAACUUU).

- 5 Furthermore, rationally designed siRNA or siRNA pools can be directed against genes involved in disorders of the brain and nervous system. Such genes would include, but are not be limited to:

APBB1

- (SEQ. ID NO. 2341: CUACGUAGCUCGUGAUAAAG,
 10 SEQ. ID NO. 2342: GCAGAGAUGUCCACACGUU,
 SEQ. ID NO. 2343: CAUGAGAUCUGCUCUAAGA,
 SEQ. ID NO. 2344: GGGCACCUCUGCUGUAUUG),

BACE1

- 15 (SEQ. ID NO. 2345: CCACAGAGCAAGUGAUUUA,
 SEQ. ID NO. 2346: GCAGAAAGGAGAUCAUUUA,
 SEQ. ID NO. 2347: GUAGCAAGAUCUUUACAUA,
 SEQ. ID NO. 2348: UGUCAGAGCUUGAUUAGAA),

20 PSEN1

- (SEQ. ID NO. 2349: GAGCUGACAUUGAAAUAUG,
 SEQ. ID NO. 2350: GUACAGCUAUUUCUCAUCA,
 SEQ. ID NO. 2351: GAGGUUAGGUGAAGUGGUU,
 SEQ. ID NO. 2352: GAAAGGGAGUCACAAGACA,
 25 SEQ. ID NO. 2353: GAACUGGAGUGGAGUAGGA,
 SEQ. ID NO. 2354: CAGCAGGCAUAUCUCAUUA,
 SEQ. ID NO. 2355: UCAAGUACCUCUCCUGAAUG),

PSEN2

- 30 (SEQ. ID NO. 2356: GCUGGGAAGUGGCUUAAUA,
 SEQ. ID NO. 2357: CAUAUUCCCUGCCCUGAUA,
 SEQ. ID NO. 2358: GGGAAGUGCUCAAGACCUA,
 SEQ. ID NO. 2359: CAUAGAAAGUGACGUGUUA),

MASS1

(SEQ. ID NO. 2360: GGAAGGAGCUGUUAUGAGA,
 SEQ. ID NO. 2361: GAAAGGAGAAGCUAAAUUA,
 SEQ. ID NO. 2362: GGAGGAAGGUCAAGAUUUA,
 5 SEQ. ID NO. 2363: GGAAAUAGCUGAGAUAAUG,),

ARX

(SEQ. ID NO. 2364: CCAGACGCCUGAUUAUUGAA,
 SEQ. ID NO. 2365: CAGCACCACUCAAGACCAA,
 10 SEQ. ID NO. 2366: CGCCUGAUUAUUGAAGUAAA,
 SEQ. ID NO. 2367: CAACAUCCACUCUCUCUUG) and

NNMT

(SEQ. ID NO. 2368: GGGCAGUGCUCAGUGGUA,
 15 SEQ. ID NO. 2369: GAAAGAGGCUGGCUACACA,
 SEQ. ID NO. 2370: GUACAGAAGUGAGACAUAA,
 SEQ. ID NO. 2371: GAGGUGAUCUCGCAAAGUU).

In addition, rationally designed siRNA or siRNA pools can be directed against
 20 genes involved in hypertension and related disorders. Such genes would include, but
 are not be limited to:

angiotensin II type 1 receptor

(SEQ. ID NO. 2372: CAAGAAGCCUGCACCAUGU,
 SEQ. ID NO. 2373: GCACUUCACUACCAAUAUGA,
 25 SEQ. ID NO. 2374: GCACUGGUCCCAAGUAGUA,
 SEQ. ID NO. 2375: CCAAAGGGCAGUAAAGUUU,
 SEQ. ID NO. 2376: GCUCAGAGGAGGUGUAUUU,
 SEQ. ID NO. 2377: GCACUUCACUACCAAUAUGA,
 SEQ. ID NO. 2378: AAAGGGCAGUAAAGUUU),

30

AGTR2

(SEQ. ID NO. 2379: GAACAUCUCUGGCAACAAU,
 SEQ. ID NO. 2380: GGUGAUUAUAUCUAAAUUG,
 SEQ. ID NO. 2381: GCAAGCAUCUUAUAUAGUU,

SEQ. ID NO. 2382: GAACCAGUCUUUCAACUCA), and other related targets.

Example XIII: Validation of Multigene Knockout using Rab5 and Eps

Two or more genes having similar, overlapping functions often leads to genetic redundancy. Mutations that knockout only one of, *e.g.*, a pair of such genes (also referred to as homologs) results in little or no phenotype due to the fact that the remaining intact gene is capable of fulfilling the role of the disrupted counterpart. To fully understand the function of such genes in cellular physiology, it is often necessary to knockout or knockdown both homologs simultaneously. Unfortunately, concomitant knockdown of two or more genes is frequently difficult to achieve in higher organisms (*e.g.* mice) thus it is necessary to introduce new technologies dissect gene function. One such approach to knocking down multiple genes simultaneously is by using siRNA. For example, **Figure 11** showed that rationally designed siRNA directed against a number of genes involved in the clathrin-mediated endocytosis pathway resulted in significant levels of protein reduction (*e.g.* >80%). To determine the effects of gene knockdown on clathrin-related endocytosis, internalization assays were performed using epidermal growth factor and transferrin. Specifically, mouse receptor-grade EGF (Collaborative Research Inc.) and iron-saturated human transferrin (Sigma) were iodinated as described previously (Jiang, X., Huang, F., Marusyk, A. & Sorkin, A. (2003) *Mol Biol Cell* 14, 858-70). HeLa cells grown in 12-well dishes were incubated with ^{125}I -EGF (1 ng/ml) or ^{125}I -transferrin (1 $\mu\text{g/ml}$) in binding medium (DMEM, 0.1% bovine serum albumin) at 37°C, and the ratio of internalized and surface radioactivity was determined during 5-min time course to calculate specific internalization rate constant k_e as described previously (Jiang, X *et al.*). The measurements of the uptakes of radiolabeled transferrin and EGF were performed using short time-course assays to avoid influence of the recycling on the uptake kinetics, and using low ligand concentration to avoid saturation of the clathrin-dependent pathway (for EGF Lund, K. A., Opresko, L. K., Starbuck, C., Walsh, B. J. & Wiley, H. S. (1990) *J. Biol. Chem.* 265, 15713-13723).

30

The effects of knocking down Rab5a, 5b, 5c, Eps, or Eps 15R (individually) are shown in **Figure 22** and demonstrate that disruption of single genes has little or no effect on EGF or Tfn internalization. In contrast, simultaneous knock down of Rab5a,

5b, and 5c, or Eps and Eps 15R, leads to a distinct phenotype (note: total concentration of siRNA in these experiments remained constant with that in experiments in which a single siRNA was introduced, see **Figure 23**). These experiments demonstrate the effectiveness of using rationally designed siRNA to

5 knockdown multiple genes and validates the utility of these reagents to override genetic redundancy.

Example XIV. Validation of Multigene Targeting Using G6PD, GAPDH, PLK, and UQC.

10 Further demonstration of the ability to knock down expression of multiple genes using rationally designed siRNA was performed using pools of siRNA directed against four separate genes. To achieve this, siRNA were transfected into cells (total siRNA concentration of 100nM) and assayed twenty-four hours later by B-DNA. Results shown in **Figure 24** show that pools of rationally designed molecules are

15 capable of simultaneously silencing four different genes.

Example XV. Validation of Multigene Knockouts As Demonstrated by Gene Expression Profiling, a Prophetic Example

20 To further demonstrate the ability to concomitantly knockdown the expression of multiple gene targets, single siRNA or siRNA pools directed against a collection of genes (*e.g.* 4, 8, 16, or 23 different targets) are simultaneously transfected into cells and cultured for twenty-four hours. Subsequently, mRNA is harvested from treated (and untreated) cells and labeled with one of two fluorescent probes dyes (*e.g.* a red fluorescent probe for the treated cells, a green fluorescent probe for the control cells.).

25 Equivalent amounts of labeled RNA from each sample is then mixed together and hybridized to sequences that have been linked to a solid support (*e.g.* a slide, "DNA CHIP"). Following hybridization, the slides are washed and analyzed to assess changes in the levels of target genes induced by siRNA.

30 Example XVI. Identifying Hyperfunctional siRNA

Identification of Hyperfunctional Bcl-2 siRNA

The ten rationally designed Bcl2 siRNA (identified in **Figure 13, 14**) were tested to identify hyperpotent reagents. To accomplish this, each of the ten Bcl-2

siRNA were individually transfected into cells at a 300pM (0.3nM) concentrations. Twenty-four hours later, transcript levels were assessed by B-DNA assays and compared with relevant controls. As shown in **Figure 25**, while the majority of Bcl-2 siRNA failed to induce functional levels of silencing at this concentration, siRNA 1 and 8 induced >80% silencing, and siRNA 6 exhibited greater than 90% silencing at this subnanomolar concentration.

By way of prophetic examples, similar assays could be performed with any of the groups of rationally designed genes described in Example VII or Example VIII.

Thus for instance, rationally designed siRNA sequences directed against PDGFA

(SEQ. ID NO. 2383: GGUAAGAUAUUGUGCUUUA,
SEQ. ID NO. 2384: CCGCAAUAUUGCAGAAUUA,
SEQ. ID NO. 2385: GGAUGUACAUGGCGUGUUA,
15 SEQ. ID NO. 2386: GGUGAAGUUUGUAUGUUUA), or

PDGFB

(SEQ. ID NO. 2387: GCUCCGCGCUUCCGAUUU,
SEQ. ID NO. 2388: GAGCAGGAAUGGUGAGAUG,
20 SEQ. ID NO. 2389: GAACUUGGGAUAAGAGUGU,
SEQ. ID NO. 2390: CCGAGGAGCUUUAUGAGAU,
SEQ. ID NO. 2391: UUUAUGAGAUGCUGAGUGA)

could be introduced into cells at increasingly limiting concentrations to determine whether any of the duplexes are hyperfunctional. Similarly, rationally designed

25 sequences directed against

HIF1 alpha

(SEQ. ID NO. 2392: GAAGGAACCUGAUGCUUUA,
SEQ. ID NO. 2393: GCAUAUAUCUAGAAGGUAU,
SEQ. ID NO. 2394: GAACAAUAACAUGGGAUUA,
30 SEQ. ID NO. 2395: GGACACAGAUUUAGACUUG), or

VEGF

(SEQ. ID NO. 2396: GAACGUACUUGCAGAUGUG,
SEQ. ID NO. 2397: GAGAAAGCAUUUGUUUGUA,

SEQ. ID NO. 2398: GGAGAAAGCAUUUGUUUGU,

SEQ. ID NO. 2399: CGAGGCAGCUUGAGUUAAA) could be introduced into cells at increasingly limiting concentrations and screened for hyperfunctional duplexes.

5 **Example XVII: Gene Silencing: Prophetic Example**

Below is an example of how one might transfect a cell.

- 10 a. Select a cell line. The selection of a cell line is usually determined by the desired application. The most important feature to RNAi is the level of expression of the gene of interest. It is highly recommended to use cell lines for which siRNA transfection conditions have been specified and validated.
- 15 b. Plate the cells. Approximately 24 hours prior to transfection, plate the cells at the appropriate density so that they will be approximately 70 – 90% confluent, or approximately 1×10^5 cells/ml at the time of transfection. Cell densities that are too low may lead to toxicity due to excess exposure and uptake of transfection reagent-siRNA complexes. Cell densities that are too high may lead to low transfection efficiencies and little or no silencing. Incubate the cells overnight. Standard incubation conditions for mammalian cells are 37°C in 5% CO₂. Other cell types, such as insect cells, require different
20 temperatures and CO₂ concentrations that are readily ascertainable by persons skilled in the art. Use conditions appropriate for the cell type of interest.
- 25 c. SiRNA re-suspension. Add 20 µl siRNA universal buffer to each siRNA to generate a final concentration of 50 µM.
- d. SiRNA-lipid complex formation. Use RNase-free solutions and tubes. Using the following table, Table XI:

e.

Table XI		
	96-well	24-well
Mixture 1 (TransIT-TKO-Plasmid dilution mixture)		
Opti-MEM	9.3 μ l	46.5 μ l
TransIT-TKO (1 μ g/ μ l)	0.5 μ l	2.5 μ l
Mixture 1 Final Volume	10.0 μl	50.0 μl
Mixture 2 (siRNA dilution mixture)		
Opti-MEM	9.0 μ l	45.0 μ l
siRNA (1 μ M)	1.0 μ l	5.0 μ l
Mixture 2 Final Volume	10.0 μl	50.0 μl
Mixture 3 (siRNA-Transfection reagent mixture)		
Mixture 1	10 μ l	50 μ l
Mixture 2	10 μ l	50 μ l
Mixture 3 Final Volume	20 μl	100 μl
Incubate 20 minutes at room temperature.		
Mixture 4 (Media-siRNA/Transfection reagent mixture)		
Mixture 3	20 μ l	100 μ l
Complete media	80 μ l	400 μ l
Mixture 4 Final Volume	100 μl	500 μl
Incubate 48 hours at 37°C.		

- 5 Transfection. Create a Mixture 1 by combining the specified amounts of OPTI-MEM serum free media and transfection reagent in a sterile polystyrene tube. Create a Mixture 2 by combining specified amounts of each siRNA with OPTI-MEM media in sterile 1 ml tubes. Create a Mixture 3 by combining specified amounts of Mixture 1 and Mixture 2. Mix gently (do not vortex) and incubate at room temperature for 20
- 10 minutes. Create a Mixture 4 by combining specified amounts of Mixture 3 to complete media. Add appropriate volume to each cell culture well. Incubate cells with transfection reagent mixture for 24 – 72 hours at 37°C. This incubation time is flexible. The ratio of silencing will remain consistent at any point in the time period. Assay for gene silencing using an appropriate detection method such as RT-PCR,
- 15 Western blot analysis, immunohistochemistry, phenotypic analysis, mass

spectrometry, fluorescence, radioactive decay, or any other method that is now known or that comes to be known to persons skilled in the art and that from reading this disclosure would be useful with the present invention. The optimal window for observing a knockdown phenotype is related to the mRNA turnover of the gene of interest, although 24 – 72 hours is standard. Final Volume reflects amount needed in each well for the desired cell culture format. When adjusting volumes for a Stock Mix, an additional 10% should be used to accommodate variability in pipetting, *etc.* Duplicate or triplicate assays should be carried out when possible.

10 While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departure from the present disclosure as come within known or customary practice
15 within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

Claims

1. A method for selecting siRNA comprising selecting an siRNA molecule of 19 – 25 nucleoside bases, said method comprising:

- (a) selecting a target gene;
- (b) measuring the functionality of sequences of 19 – 25 nucleotides in length that are substantially complementary to a stretch of nucleotides of the target sequence, wherein said functionality is dependent upon non-target specific criteria.

2. The method according to claim 1 wherein said functionality is determined by applying one of the following formulas:

$$\text{Formula I} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula II} = -(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$$

$$\text{Formula III} = -(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3;$$

$$\text{Formula IV} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula V} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$$

$$\text{Formula VI} = -(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$$

$$\text{Formula VII} = -(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3 + (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2;$$

wherein in Formulas I – VII:

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense

strand at positions 15 –19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$

wherein

- $A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
5 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
10 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

- $C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
15 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
20 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

- 25 $G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;
30 $G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base
is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;
 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;

- $U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;
 $U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;
 $U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;
 $U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;
5 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;
 $U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;
 $U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;
 $U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;
 $U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;
10 GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand
or within positions 15 –18 if the sense strand is only 18 base pairs in length;
 GC_{total} = the number of G and C bases in the sense strand;
 $T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs,
15 otherwise its value is 0; and
 X = the number of times that the same nucleotide repeats four or more times in a
row.

3. A method of gene-silencing comprising selecting an siRNA according to
20 claim 2 and introducing it into a cell.
4. The method according to claim 3 wherein said introducing is by allowing
passive uptake of the siRNA.
- 25 5. The method according to claim 3, wherein said introducing is through the
use of a vector.
6. A method for developing an siRNA algorithm for selecting siRNA, said
method comprising:
30 (a) selecting a set of siRNA;
(b) measuring the gene silencing ability of each siRNA from said set;
(c) determining the relative functionality of each siRNA;
(d) determining the amount of improved functionality by the presence or
absence of at least one variable selected from the group consisting of

the total GC content, melting temperature of the siRNA, GC content at positions 15 –19, the presence or absence of a particular nucleotide at a particular position and the number of times that the same nucleotide repeats within a given sequence; and

- 5 (e) developing an algorithm using the information of step (d).
7. A method of selecting an siRNA with improved functionality, said method comprising using the algorithm of claim 6.
- 10 8. A method of selecting hyperfunctional siRNA, said method comprising using at least one functional siRNA, wherein at least one said functional siRNA has been selected according to the method of claim 7 and measuring the silencing ability of said at least one functional siRNA, wherein silencing ability is measured at a concentration of less than 1
15 nanomolar siRNA.
9. An siRNA molecule, wherein said siRNA molecule is effective at silencing Bcl-2.
- 20 10. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a sequence substantially similar to a sequence selected from the group consisting of GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301);
GAAGUACAUC CAUUAUAAG (SEQ. ID NO. 302);
GUACGACAACCGGGAGUA (SEQ. ID NO. 303);
25 AGAUAGUGAUGAAGUACAU (SEQ. ID NO. 304);
UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305);
CAUGC GCCCUCUGUUUGA (SEQ. ID NO. 306);
UGCGGCCCUCUGUUUGAUUU (SEQ. ID NO. 307);
GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);
30 GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and
GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).
11. The siRNA molecule of claim 10, wherein said siRNA molecule comprises a sequence selected from the group consisting of

GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301);
 GAAGUACAUCCAUUAUAAG (SEQ. ID NO. 302);
 GUACGACAACCGGGAGUA (SEQ. ID NO. 303);
 AGAUAGUGAUGAAGUACA (SEQ. ID NO. 304);
 5 UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305);
 GCAUGCGGCCUCUGUUUGA (SEQ. ID NO. 306);
 UGCGGCCUCUGUUUGAUUU (SEQ. ID NO. 307);
 GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);
 GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and
 10 GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).

12. The siRNA molecule of claim 11, wherein said siRNA molecule comprises GCAUGCGGCCUCUGUUUGA .
- 15 13. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a sense strand and an anti-sense strand.
14. The siRNA molecule of claim 9, wherein said siRNA molecule comprises a hairpin.
- 20 15. The siRNA molecule of claim 9, wherein said siRNA molecule comprises between 18 and 30 base pairs.
- 25 16. A kit for gene silencing comprising at least one siRNA selected from the group consisting of sequences substantially similar to the group consisting of GGGAGAUAGUGAUGAAGUA (SEQ. ID NO. 301);
 GAAGUACAUCCAUUAUAAG (SEQ. ID NO. 302);
 GUACGACAACCGGGAGUA (SEQ. ID NO. 303);
 AGAUAGUGAUGAAGUACA (SEQ. ID NO. 304);
 30 UGAAGACUCUGCUCAGUUU (SEQ. ID NO. 305);
 GCAUGCGGCCUCUGUUUGA (SEQ. ID NO. 306);
 UGCGGCCUCUGUUUGAUUU (SEQ. ID NO. 307);
 GAGAUAGUGAUGAAGUACA (SEQ. ID NO. 308);

GGAGAUAGUGAUGAAGUAC (SEQ. ID NO. 309); and
GAAGACUCUGCUCAGUUUG (SEQ. ID NO. 310).

17. A method of gene silencing comprising using the siRNA molecule of
5 claim 10.
18. A method of gene silencing comprising using the siRNA molecule of claim
11.
- 10 19. A kit, wherein said kit is comprised of at least two siRNA, wherein said at
least two siRNA comprise a first optimized siRNA and a second optimized
siRNA, wherein said first optimized siRNA and said second optimized
siRNA are optimized according to one of the following formulas:
- 15 Formula I = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3)$
 $+ (U_{10}) + (A_{14}) - (U_5) - (A_{11});$
- Formula II = $-(GC/3) - (AU_{15-19}) * 3 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$
- 20 Formula III = $-(GC/3) + (AU_{15-19}) - (Tm_{20^{\circ}C}) * 3;$
- Formula IV = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 2 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 2$
 $+ (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$
- 25 Formula V = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3) + (U_{10}) + (A_{14}) - (U_5) - (A_{11});$
- Formula VI = $-(G_{13}) * 3 - (C_{19}) + (A_{19}) * 2 + (A_3);$
- Formula VII = $-(GC/2) + (AU_{15-19})/2 - (Tm_{20^{\circ}C}) * 1 - (G_{13}) * 3 - (C_{19}) + (A_{19}) * 3$
30 $+ (A_3) * 3 + (U_{10})/2 + (A_{14})/2 - (U_5)/2 - (A_{11})/2;$

wherein in Formulas I – VII:

$AU_{15-19} = 0 - 5$ depending on the number of A or U bases on the sense strand at positions 15 -19;

$G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise its value is 0;

GC = the number of G and C bases in the entire sense strand;

$Tm_{20^{\circ}C} = 1$ if the Tm is greater than $20^{\circ}C$;

$A_3 = 1$ if A is the base at position 3 on the sense strand, otherwise its value is 0;

$A_{11} = 1$ if A is the base at position 11 on the sense strand, otherwise its value is 0;

$A_{14} = 1$ if A is the base at position 14 on the sense strand, otherwise its value is 0;

$A_{19} = 1$ if A is the base at position 19 on the sense strand, otherwise its value is 0;

$U_5 = 1$ if U is the base at position 5 on the sense strand, otherwise its value is 0;

$U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

or,

Formula VIII: $(-14)*G_{13}-13*A_1-12*U_7-11*U_2-10*A_{11}-10*U_4-10*C_3-10*C_5-10*C_6-9*A_{10}-9*U_9-9*C_{18}-8*G_{10}-7*U_1-7*U_{16}-7*C_{17}-7*C_{19}+7*U_{17}+8*A_2+8*A_4+8*A_5+8*C_4+9*G_8+10*A_7+10*U_{18}+11*A_{19}+11*C_9+15*G_1+18*A_3+19*U_{10}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$; and

Formula IX: $(14.1)*A_3+(14.9)*A_6+(17.6)*A_{13}+(24.7)*A_{19}+(14.2)*U_{10}+(10.5)*C_9+(23.9)*G_1+(16.3)*G_2+(-12.3)*A_{11}+(-19.3)*U_1+(-12.1)*U_2+$

$$(-11)*U_3+(-15.2)*U_{15}+(-11.3)*U_{16}+(-11.8)*C_3+(-17.4)*C_6+(-10.5)*C_7+(-13.7)*G_{13}+(-25.9)*G_{19}-Tm-3*(GC_{total})-6*(GC_{15-19})-30*X$$

wherein

- 5 $A_1 = 1$ if A is the base at position 1 of the sense strand, otherwise its value is 0;
 $A_2 = 1$ if A is the base at position 2 of the sense strand, otherwise its value is 0;
 $A_3 = 1$ if A is the base at position 3 of the sense strand, otherwise its value is 0;
 $A_4 = 1$ if A is the base at position 4 of the sense strand, otherwise its value is 0;
 $A_5 = 1$ if A is the base at position 5 of the sense strand, otherwise its value is 0;
10 $A_6 = 1$ if A is the base at position 6 of the sense strand, otherwise its value is 0;
 $A_7 = 1$ if A is the base at position 7 of the sense strand, otherwise its value is 0;
 $A_{10} = 1$ if A is the base at position 10 of the sense strand, otherwise its value is 0;
 $A_{11} = 1$ if A is the base at position 11 of the sense strand, otherwise its value is 0;
 $A_{13} = 1$ if A is the base at position 13 of the sense strand, otherwise its value is 0;
15 $A_{19} = 1$ if A is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;
- $C_3 = 1$ if C is the base at position 3 of the sense strand, otherwise its value is 0;
 $C_4 = 1$ if C is the base at position 4 of the sense strand, otherwise its value is 0;
20 $C_5 = 1$ if C is the base at position 5 of the sense strand, otherwise its value is 0;
 $C_6 = 1$ if C is the base at position 6 of the sense strand, otherwise its value is 0;
 $C_7 = 1$ if C is the base at position 7 of the sense strand, otherwise its value is 0;
 $C_9 = 1$ if C is the base at position 9 of the sense strand, otherwise its value is 0;
 $C_{17} = 1$ if C is the base at position 17 of the sense strand, otherwise its value is 0;
25 $C_{18} = 1$ if C is the base at position 18 of the sense strand, otherwise its value is 0;
 $C_{19} = 1$ if C is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;
- $G_1 = 1$ if G is the base at position 1 on the sense strand, otherwise its value is 0;
30 $G_2 = 1$ if G is the base at position 2 of the sense strand, otherwise its value is 0;
 $G_8 = 1$ if G is the base at position 8 on the sense strand, otherwise its value is 0;
 $G_{10} = 1$ if G is the base at position 10 on the sense strand, otherwise its value is 0;
 $G_{13} = 1$ if G is the base at position 13 on the sense strand, otherwise its value is 0;

$G_{19} = 1$ if G is the base at position 19 of the sense strand, otherwise if another base is present or the sense strand is only 18 base pairs in length, its value is 0;

$U_1 = 1$ if U is the base at position 1 on the sense strand, otherwise its value is 0;

5 $U_2 = 1$ if U is the base at position 2 on the sense strand, otherwise its value is 0;

$U_3 = 1$ if U is the base at position 3 on the sense strand, otherwise its value is 0;

$U_4 = 1$ if U is the base at position 4 on the sense strand, otherwise its value is 0;

$U_7 = 1$ if U is the base at position 7 on the sense strand, otherwise its value is 0;

$U_9 = 1$ if U is the base at position 9 on the sense strand, otherwise its value is 0;

10 $U_{10} = 1$ if U is the base at position 10 on the sense strand, otherwise its value is 0;

$U_{15} = 1$ if U is the base at position 15 on the sense strand, otherwise its value is 0;

$U_{16} = 1$ if U is the base at position 16 on the sense strand, otherwise its value is 0;

$U_{17} = 1$ if U is the base at position 17 on the sense strand, otherwise its value is 0;

$U_{18} = 1$ if U is the base at position 18 on the sense strand, otherwise its value is 0;

15

GC_{15-19} = the number of G and C bases within positions 15 – 19 of the sense strand
or within positions 15 –18 if the sense strand is only 18 base pairs in length;

GC_{total} = the number of G and C bases in the sense strand;

$T_m = 100$ if the targeting site contains an inverted repeat longer than 4 base pairs,

20 otherwise its value is 0; and

X = the number of times that the same nucleotide repeats four or more times in a row.

Figure 1

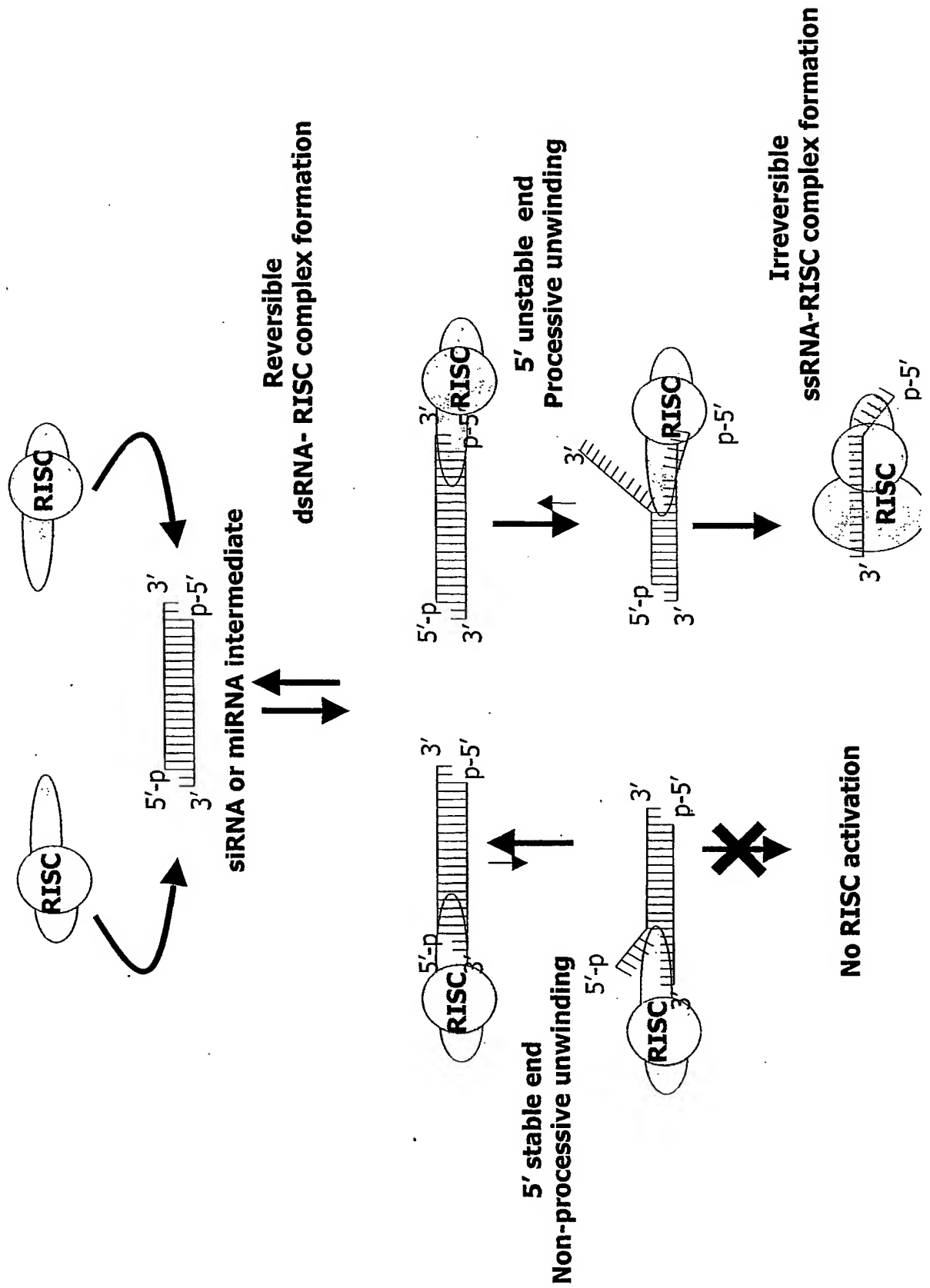


Figure 2

siRNA panel (270)

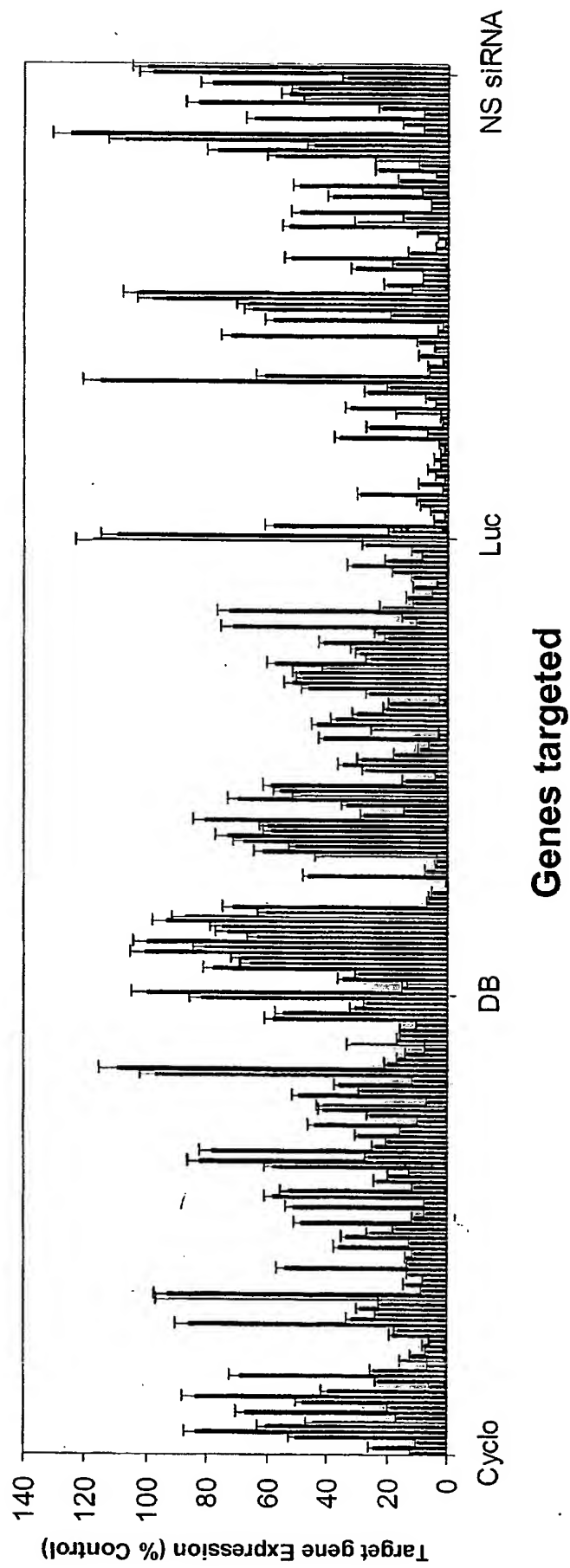
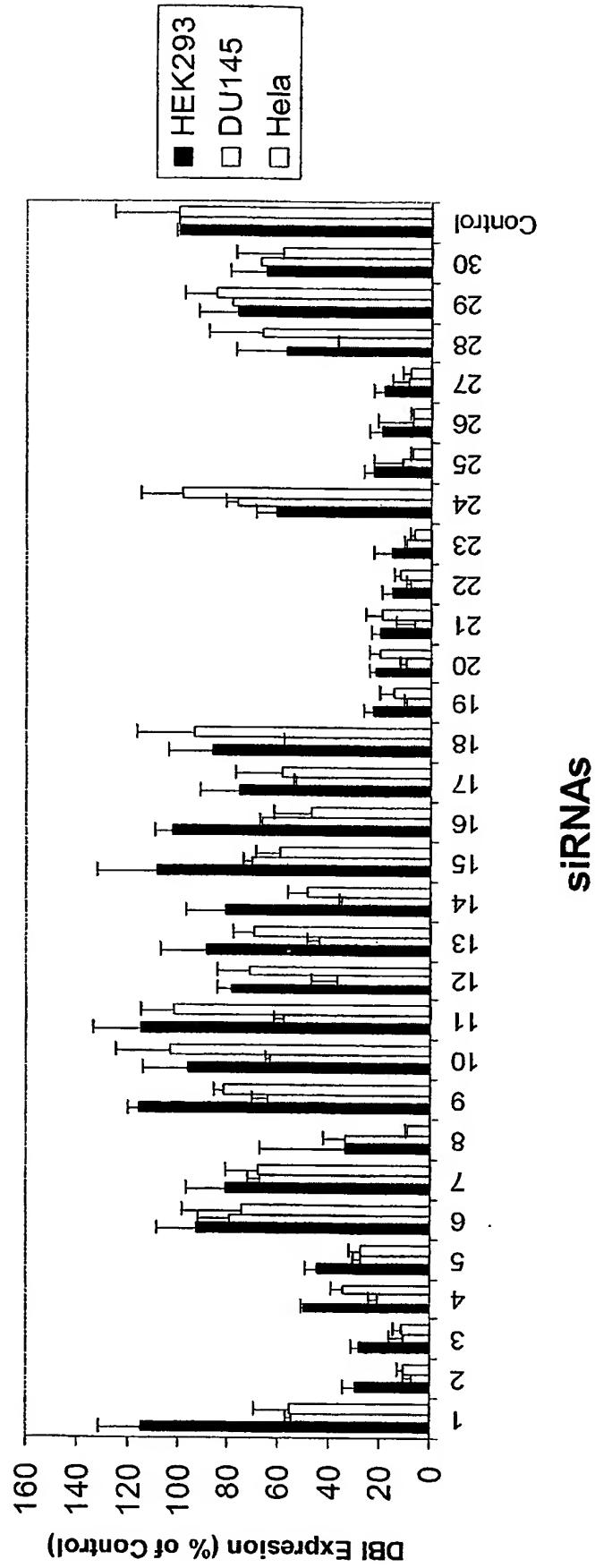


Figure 3a

siRNA functionality is independent from the cell line



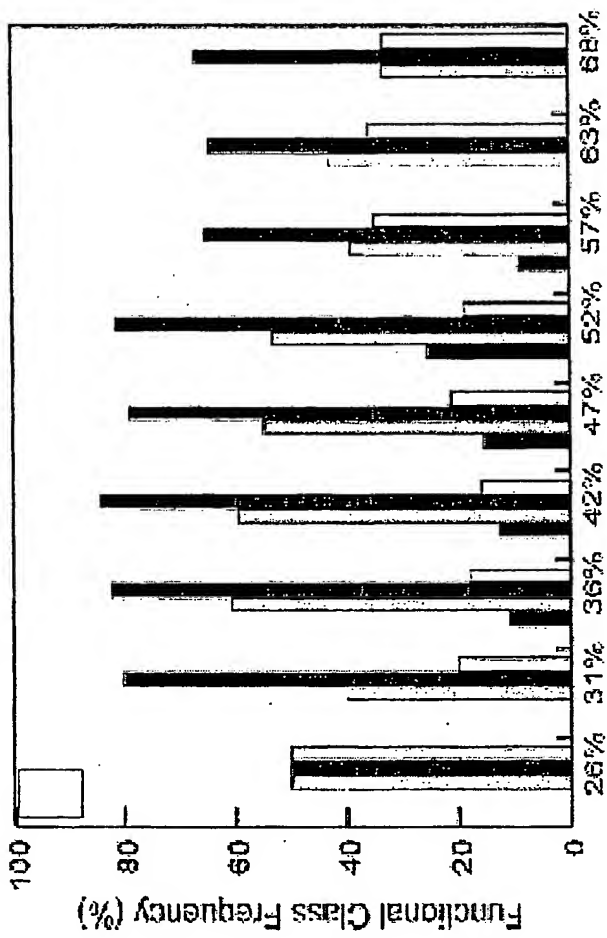


Figure 3b

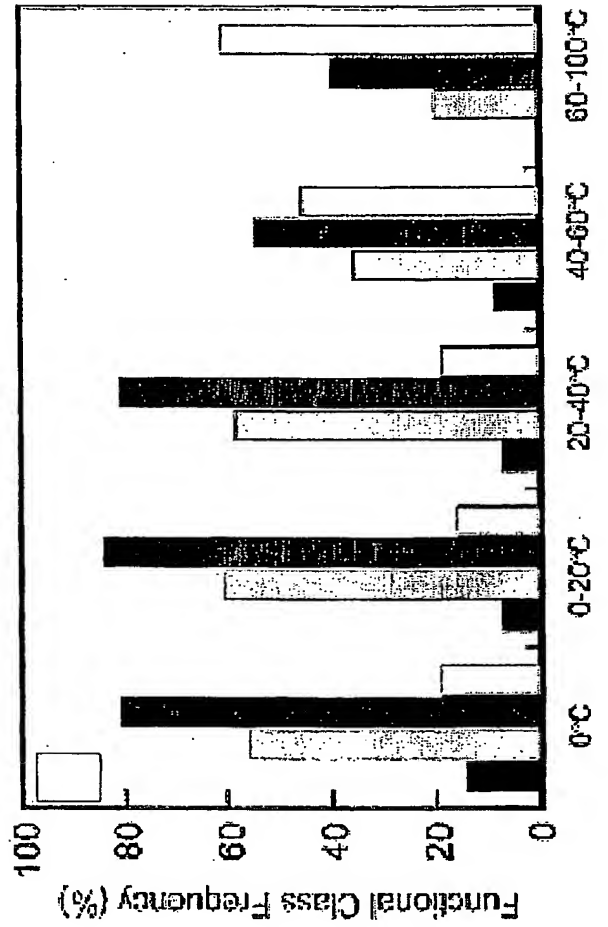


Figure 3c

Figure 4

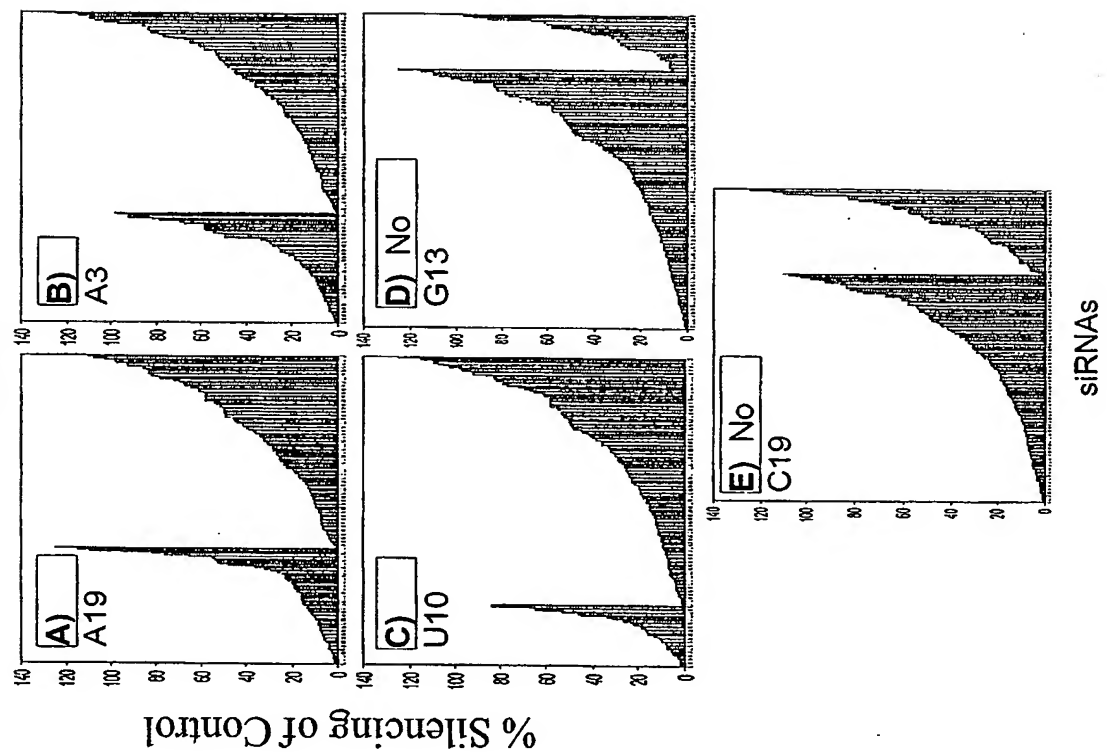


Figure 5A

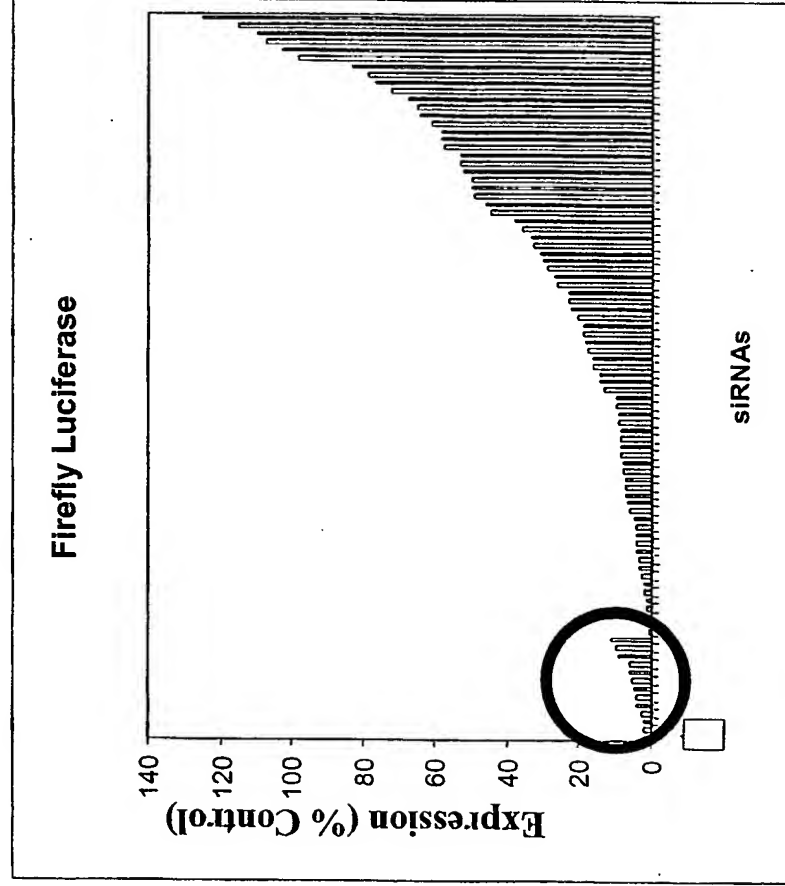


Figure 5B

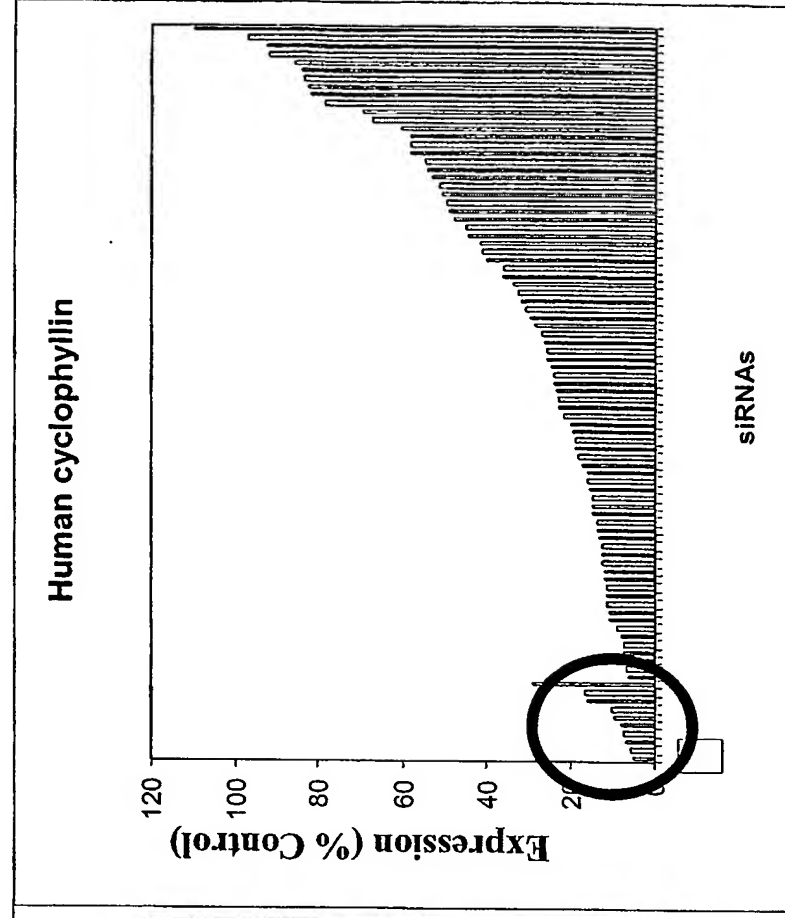


Figure 6a

Differential internal stability

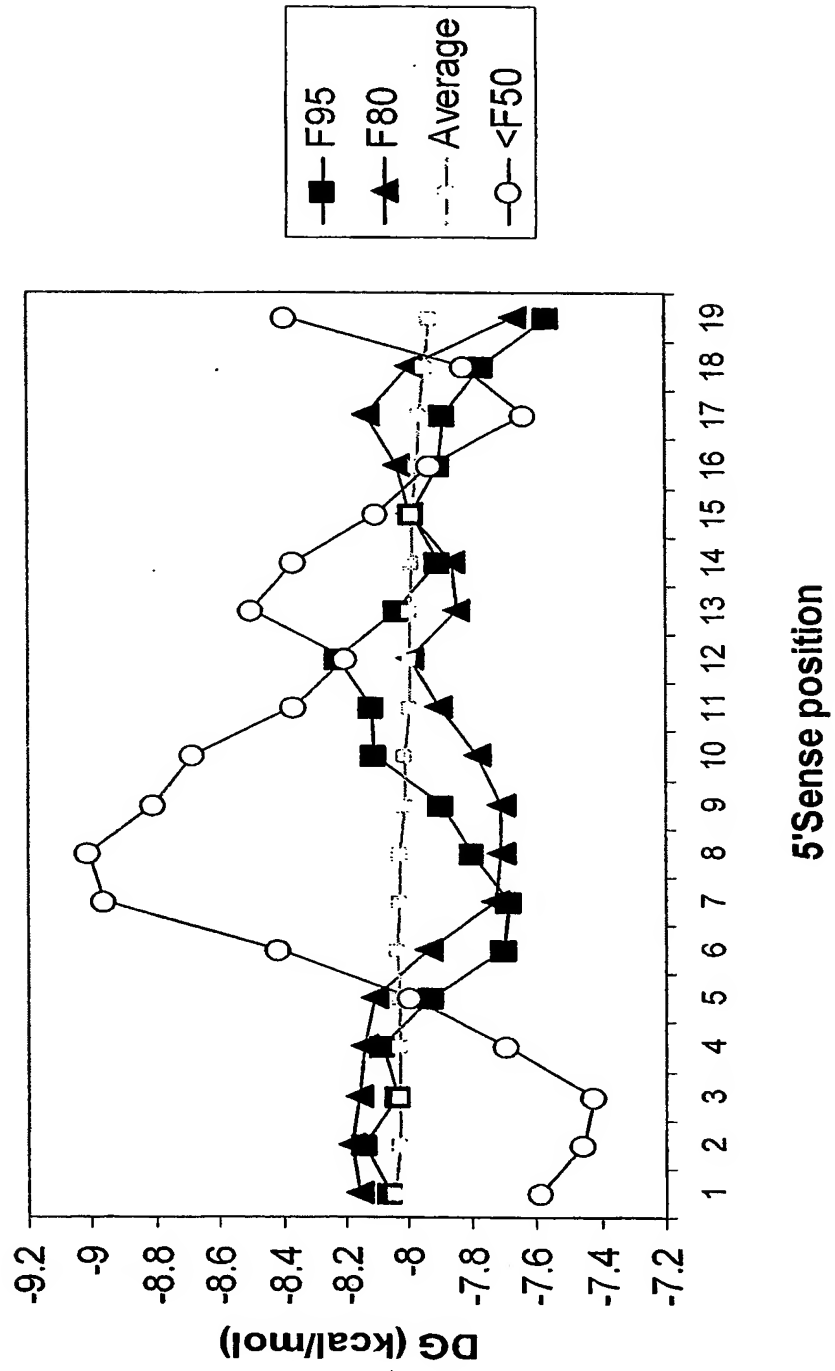


Figure 6b

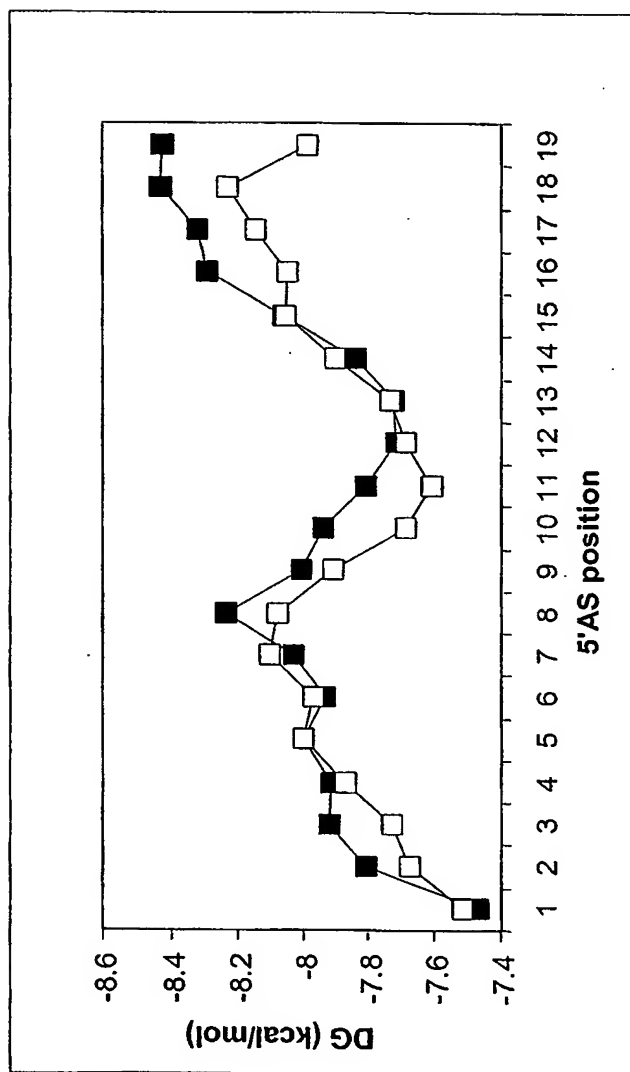
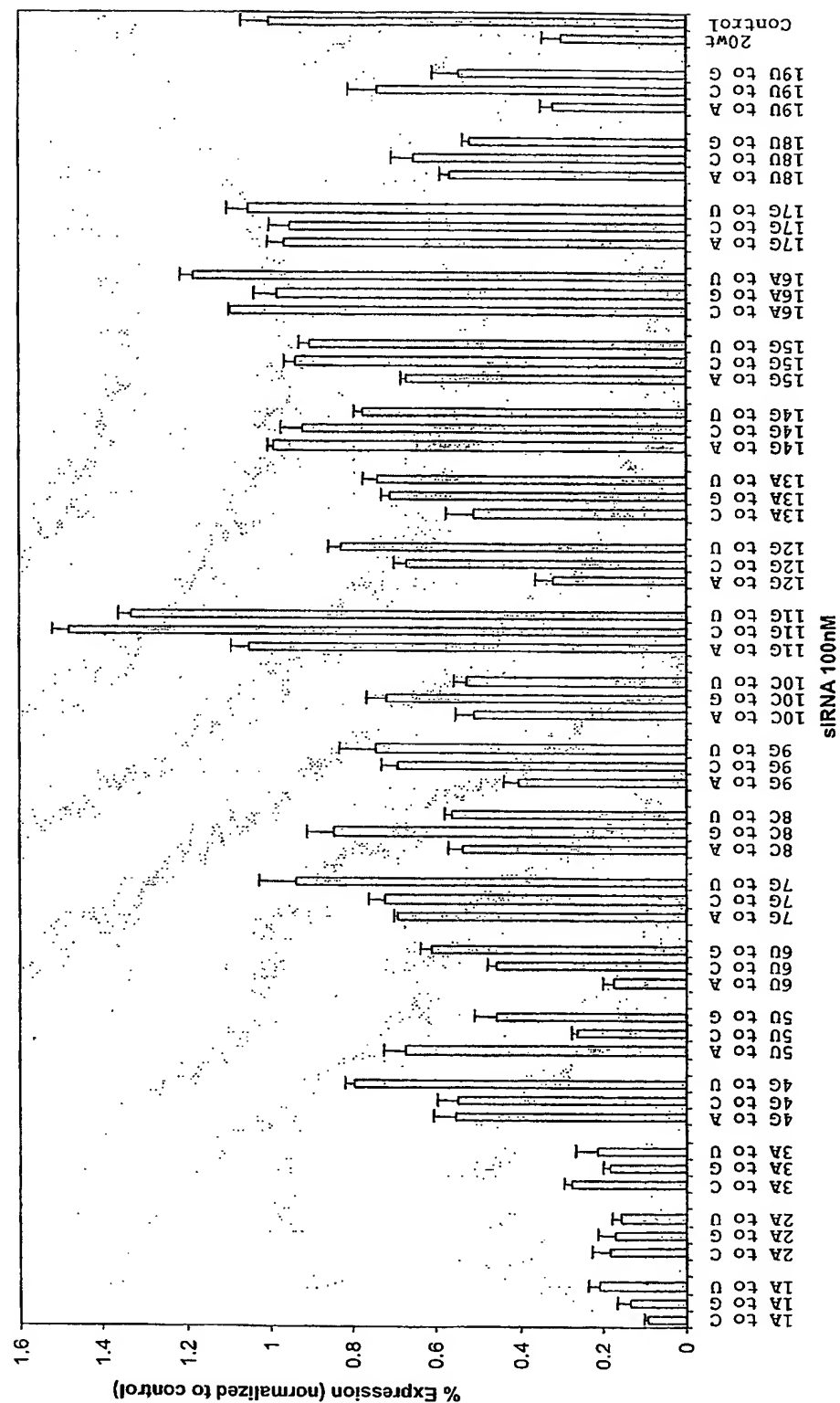


Figure 7



TARGET Screen Normalized LUC ASSAY 293 cells

Figure 8a

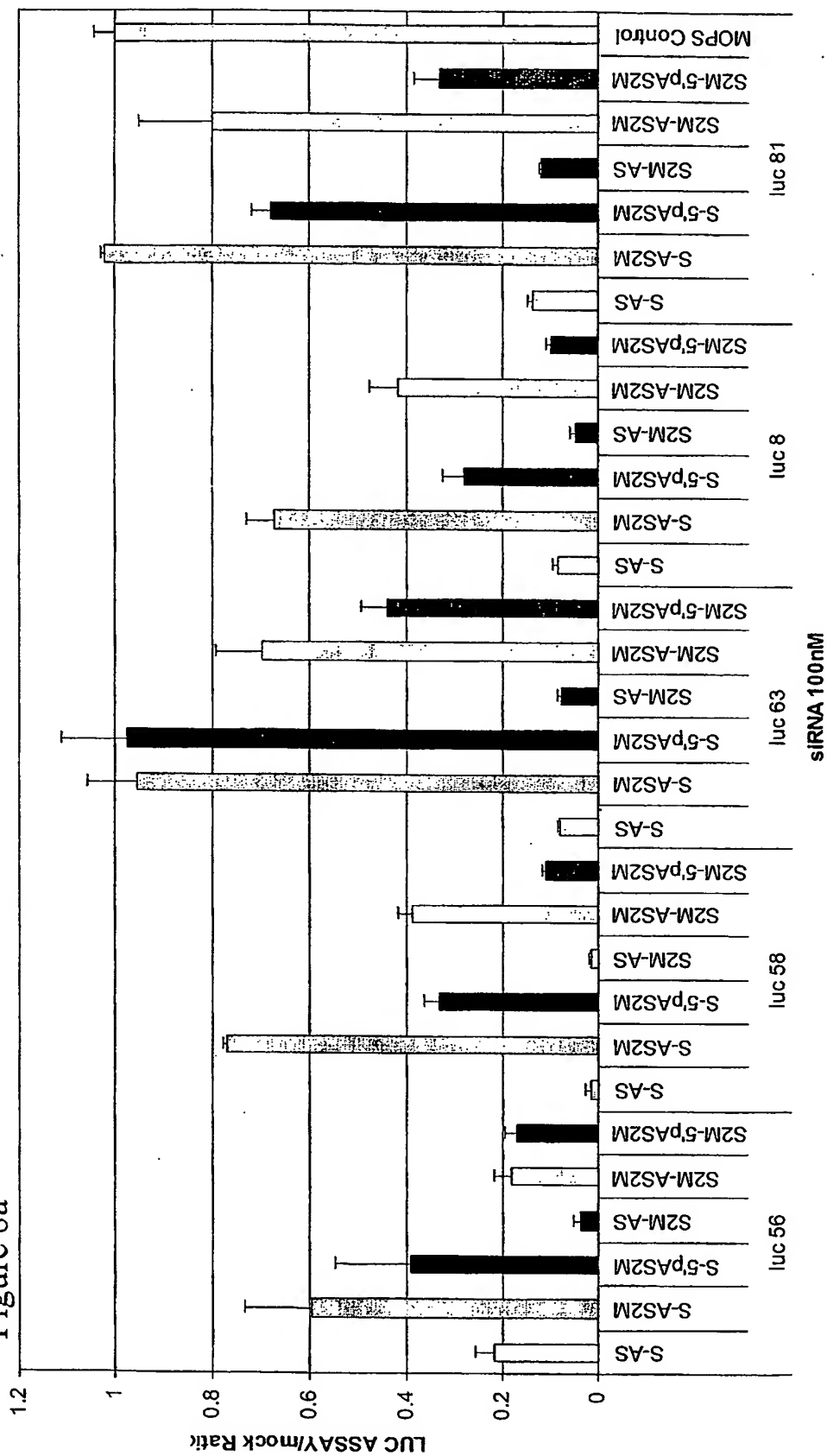


Figure 9

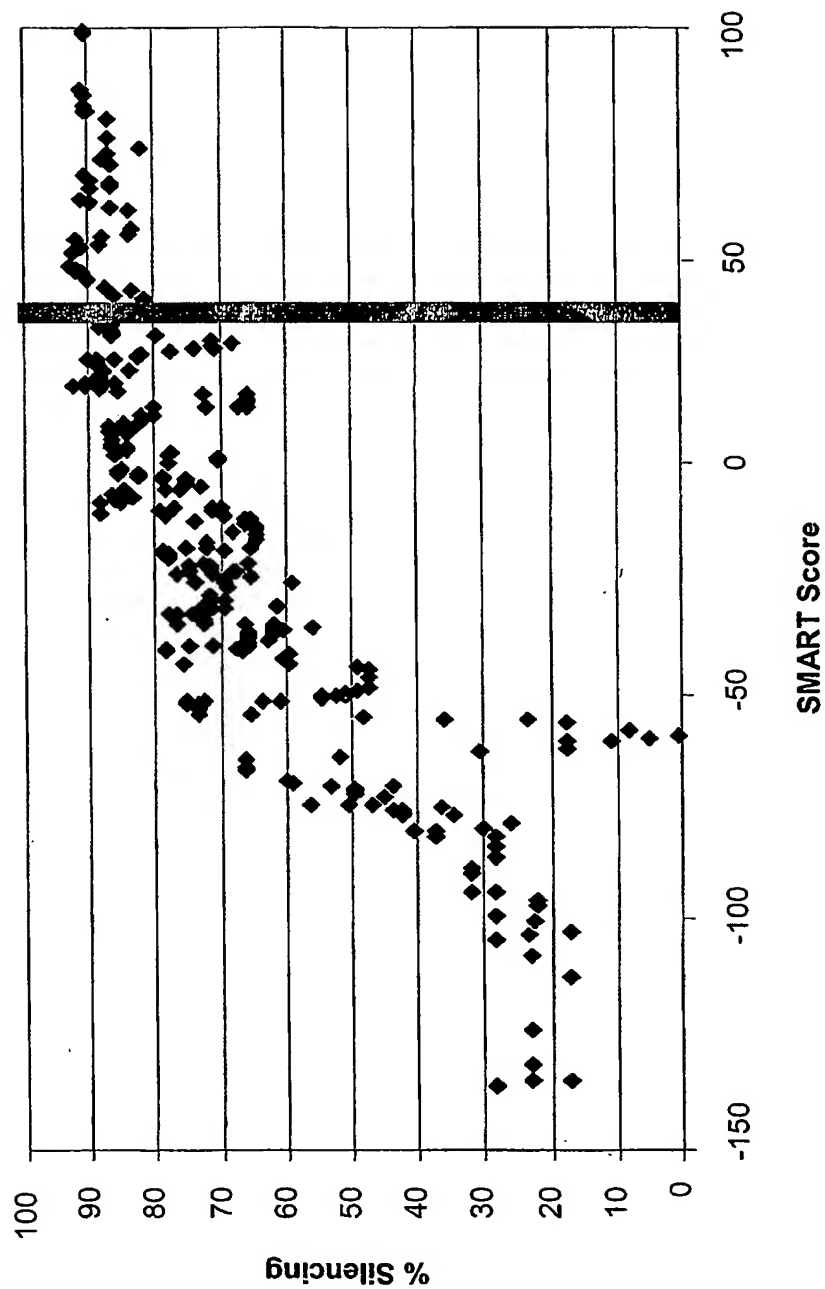
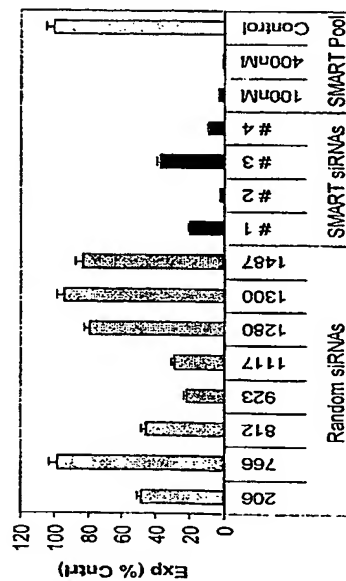
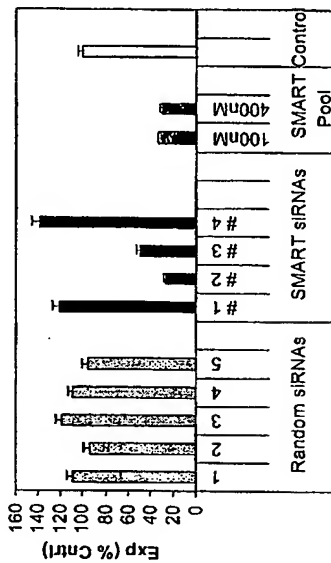


Figure 10a-f

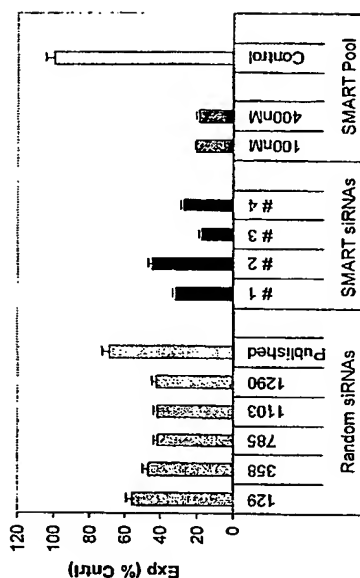
a. Human Secreted Alkaline Phosphatase



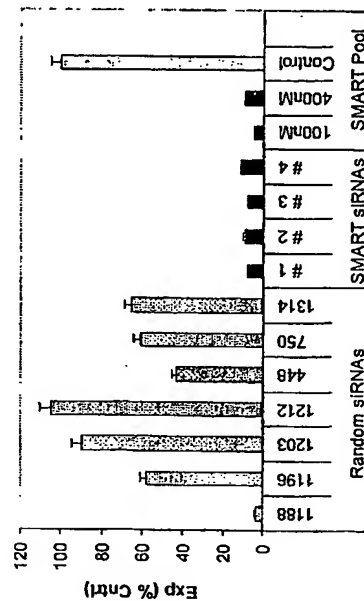
b. Homo sapiens Acyl-Coenzyme A binding protein (DBI)



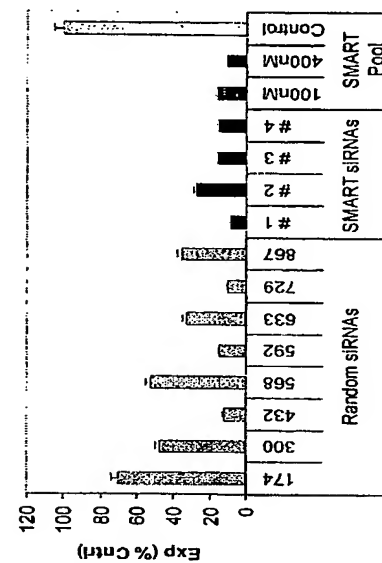
c. Homo sapiens polo-like kinase (PLK)



d. Firefly Luciferase



e. Renilla Luciferase



f. EGFR

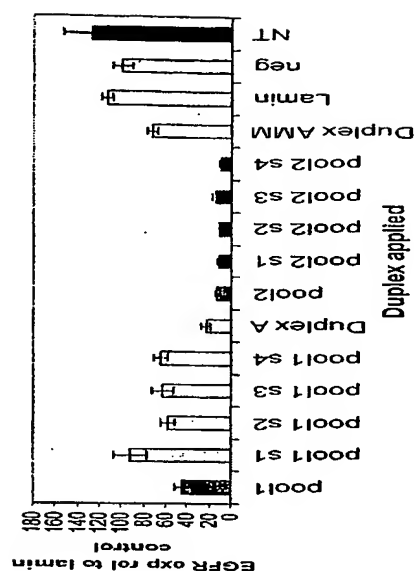


Figure 11

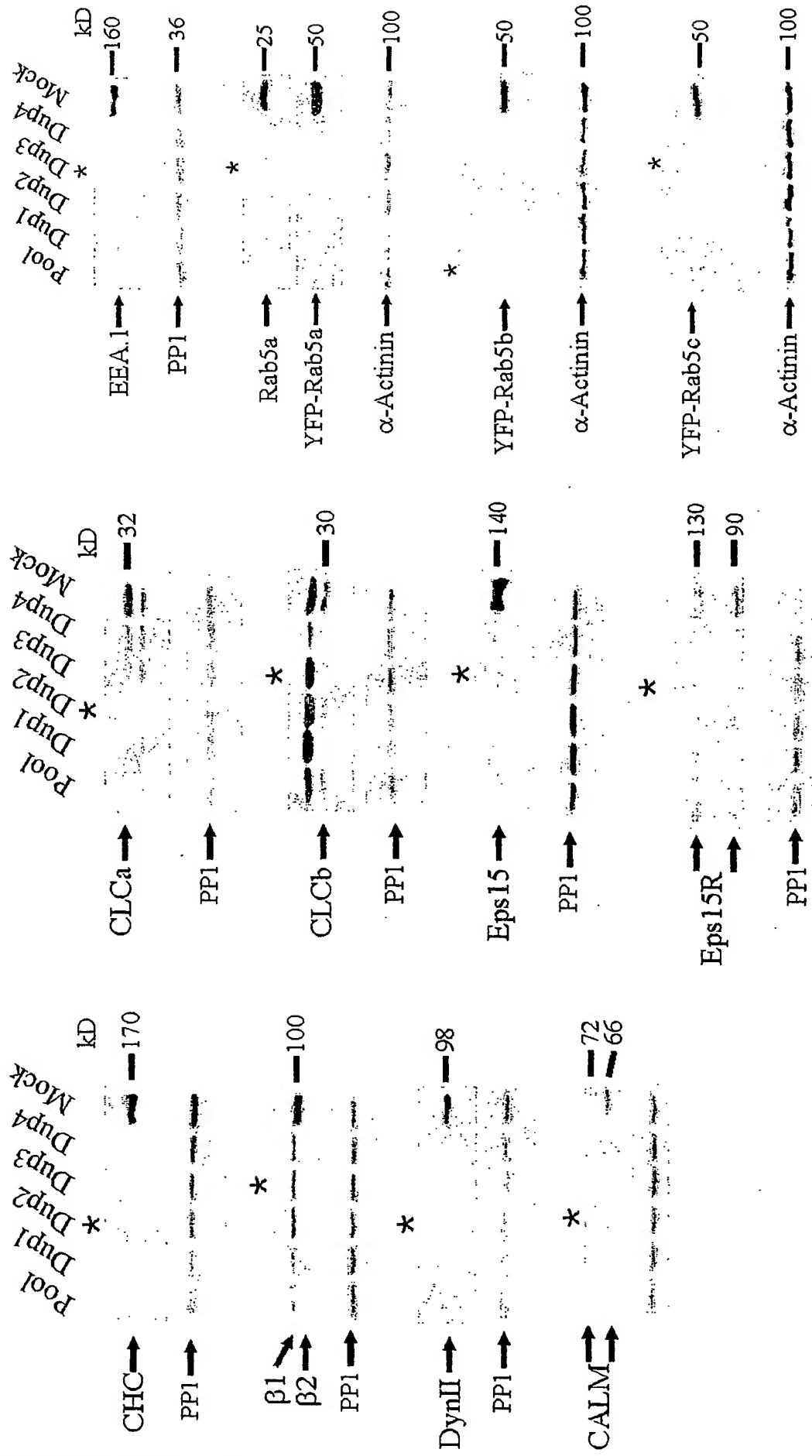


Figure 12

Rational selection validation

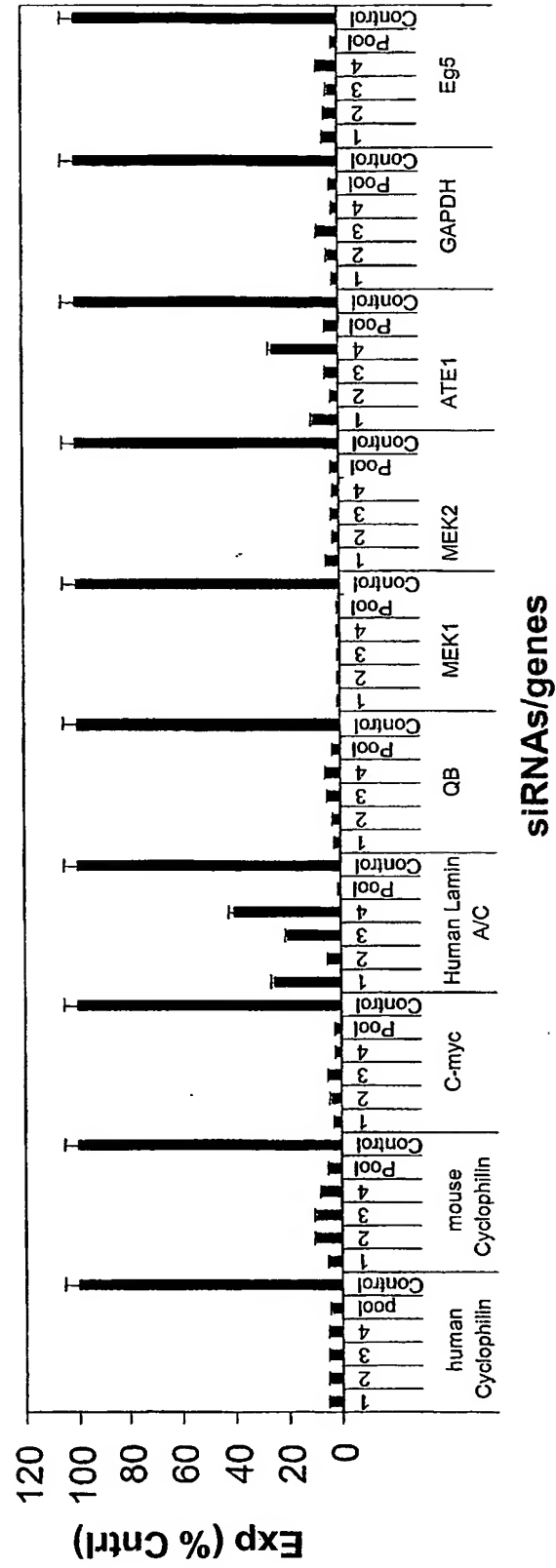
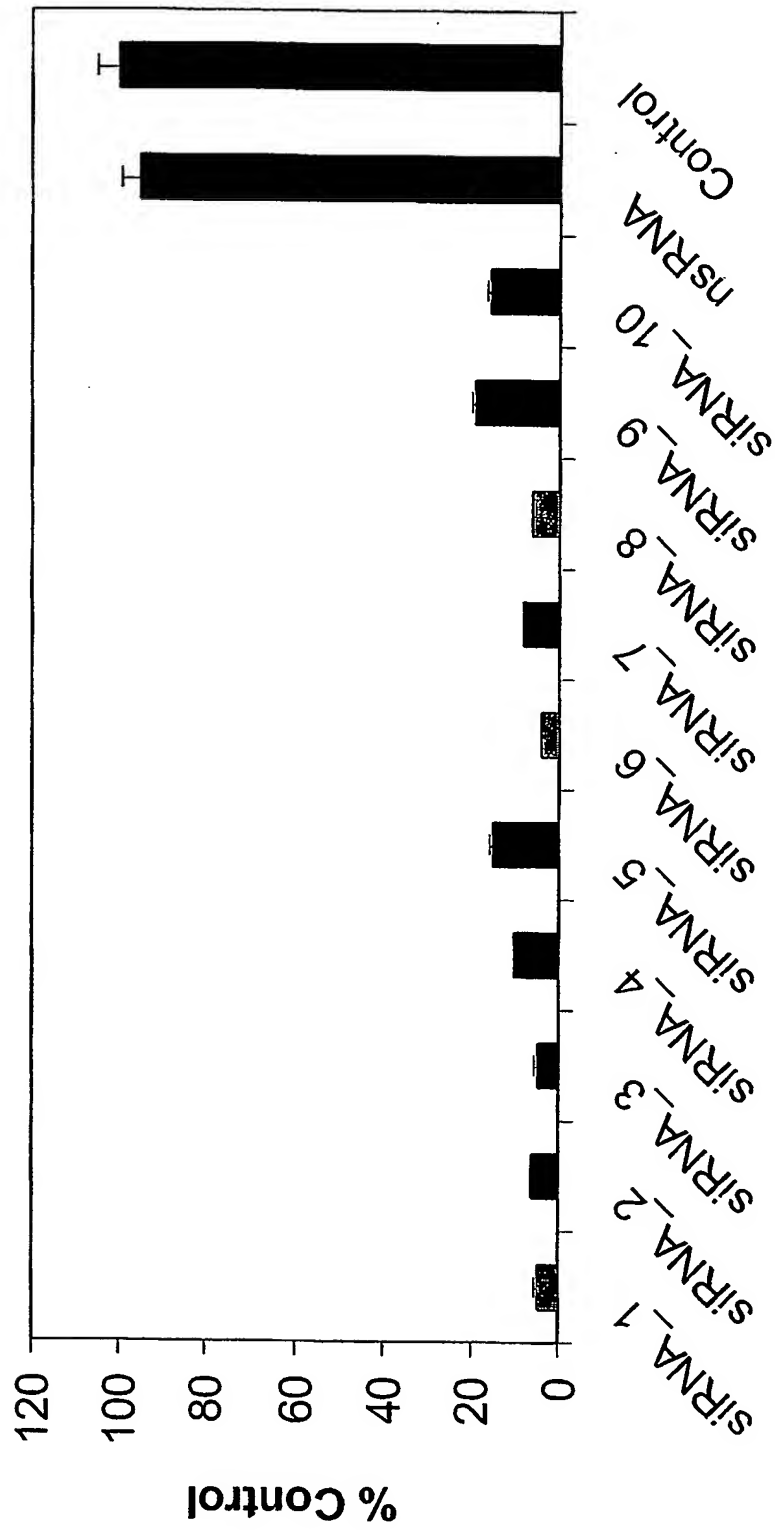


Figure 13 Sequences of top Bcl2

siRNA 1 GGGAGAUAGUGAAGUA
siRNA 2 GAAGUACAUCCAUAUAAG
siRNA 3 GUACGACAACCGGAGUA
siRNA 4 AGAUAGUGAUGAAGUACAU
siRNA 5 UGAAGACUCUGCUCAGUUU
siRNA 6 GCAUGCGGCCUCUGUUUGA
siRNA 7 UGCGGCCUCUGUUUGAUUU
siRNA 8 GAGUAAGUGAUGAAGUACA
siRNA 9 GGAGUAAGUGAUGAAGUAC
siRNA 10 GAAGACUCUGCUCAGUUUG

Figure 14

**Bcl-2 knockdown by 10 rationally designed siRNAs at
100 nM concentration**



Reporter gene individual siRNAs walk

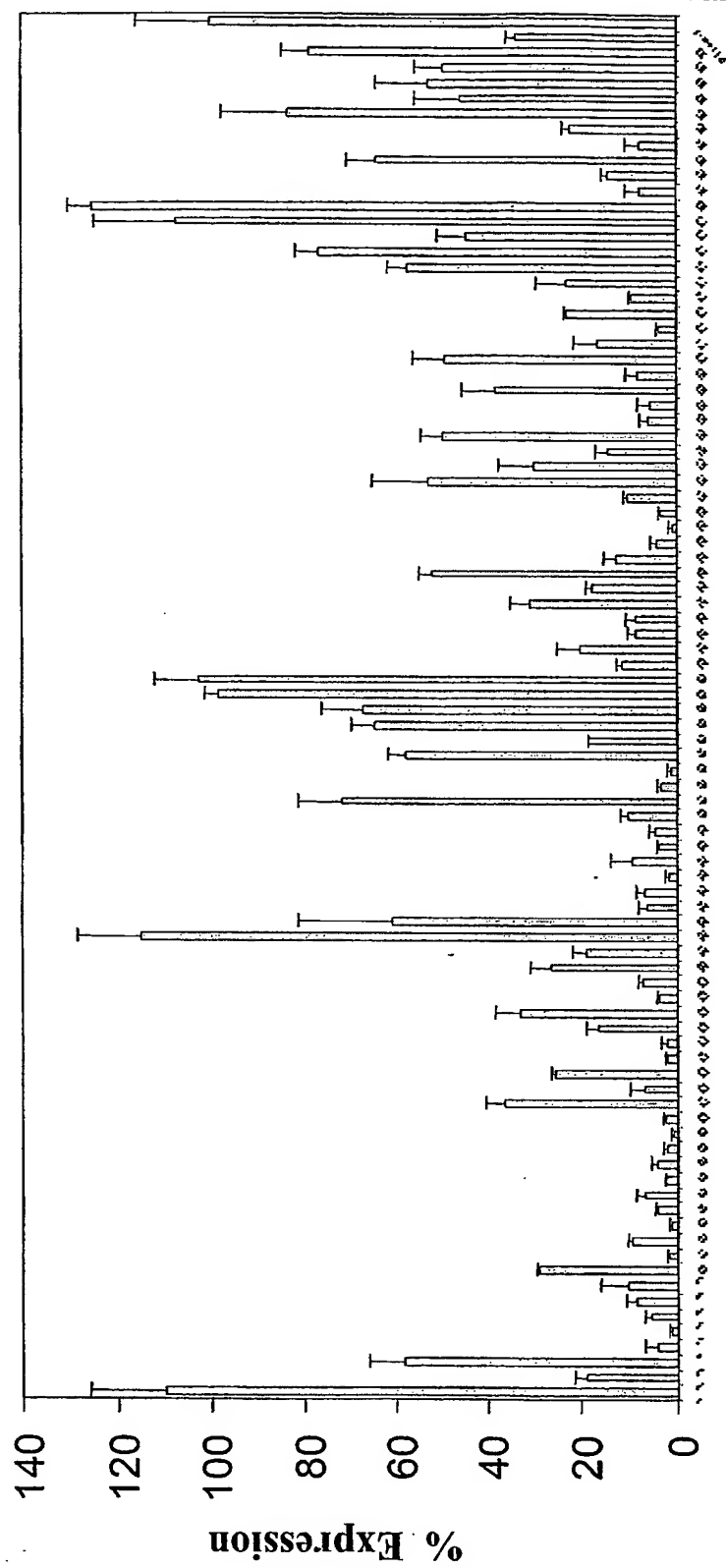


Figure 15

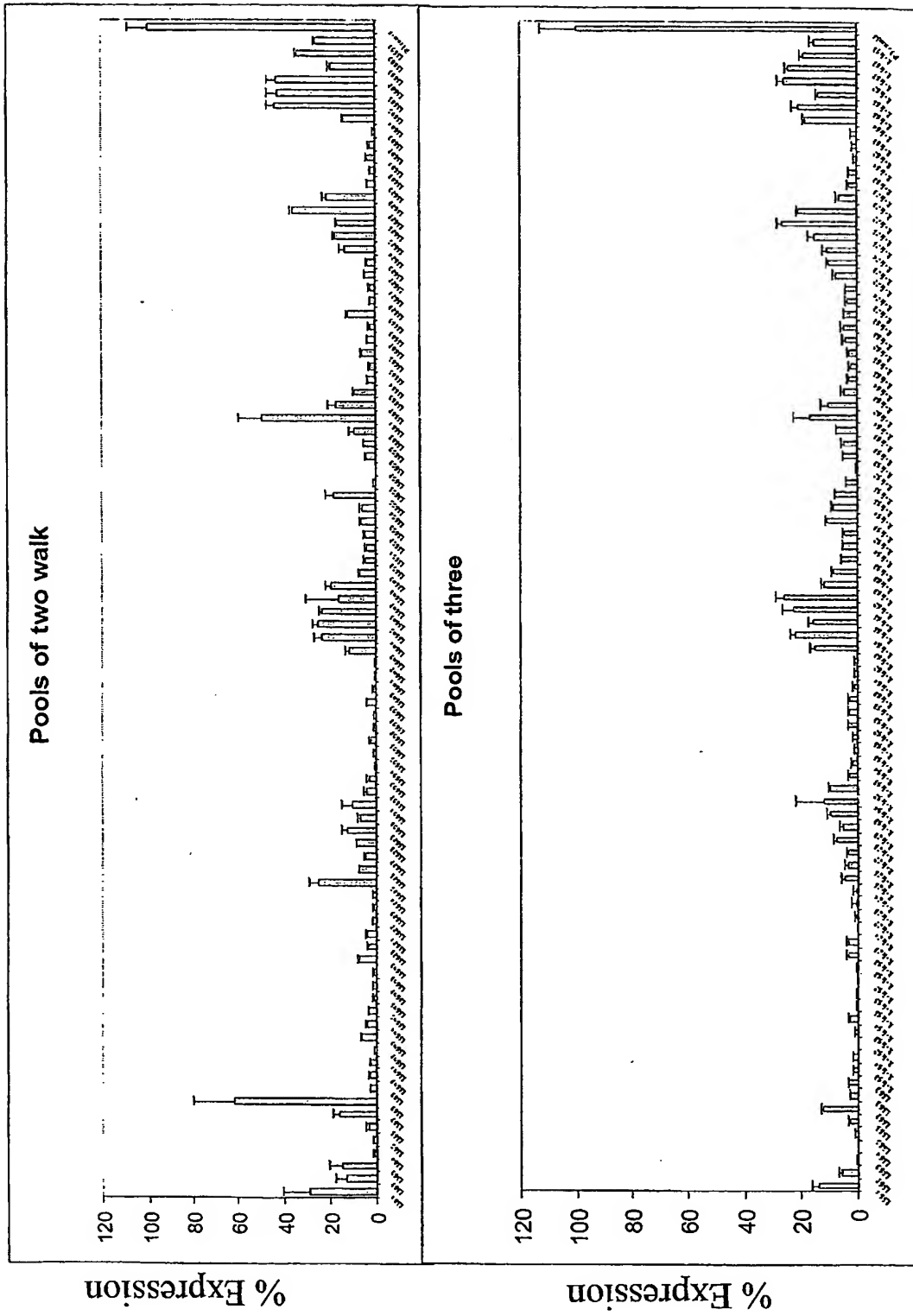


Figure 16

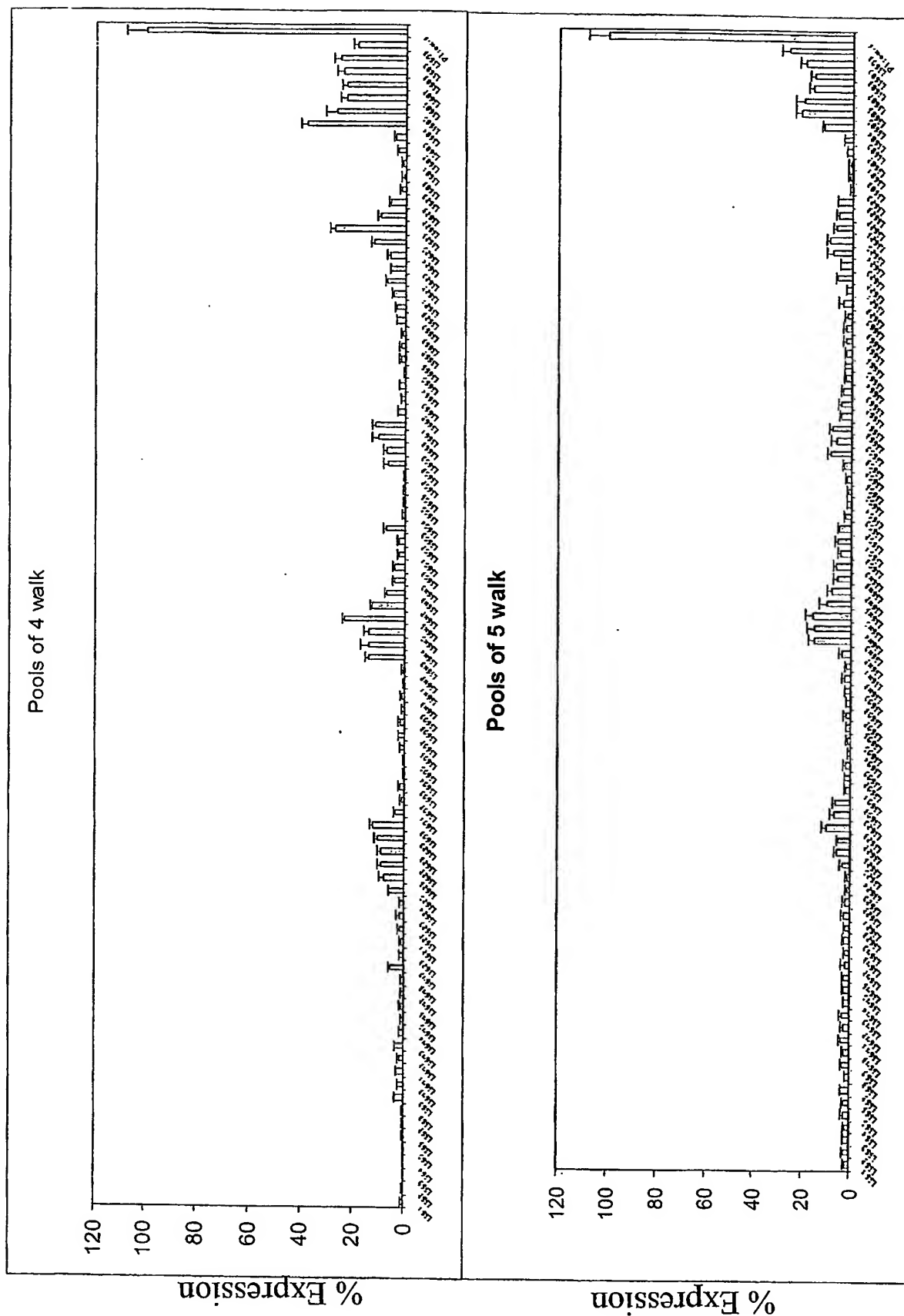


Figure 17

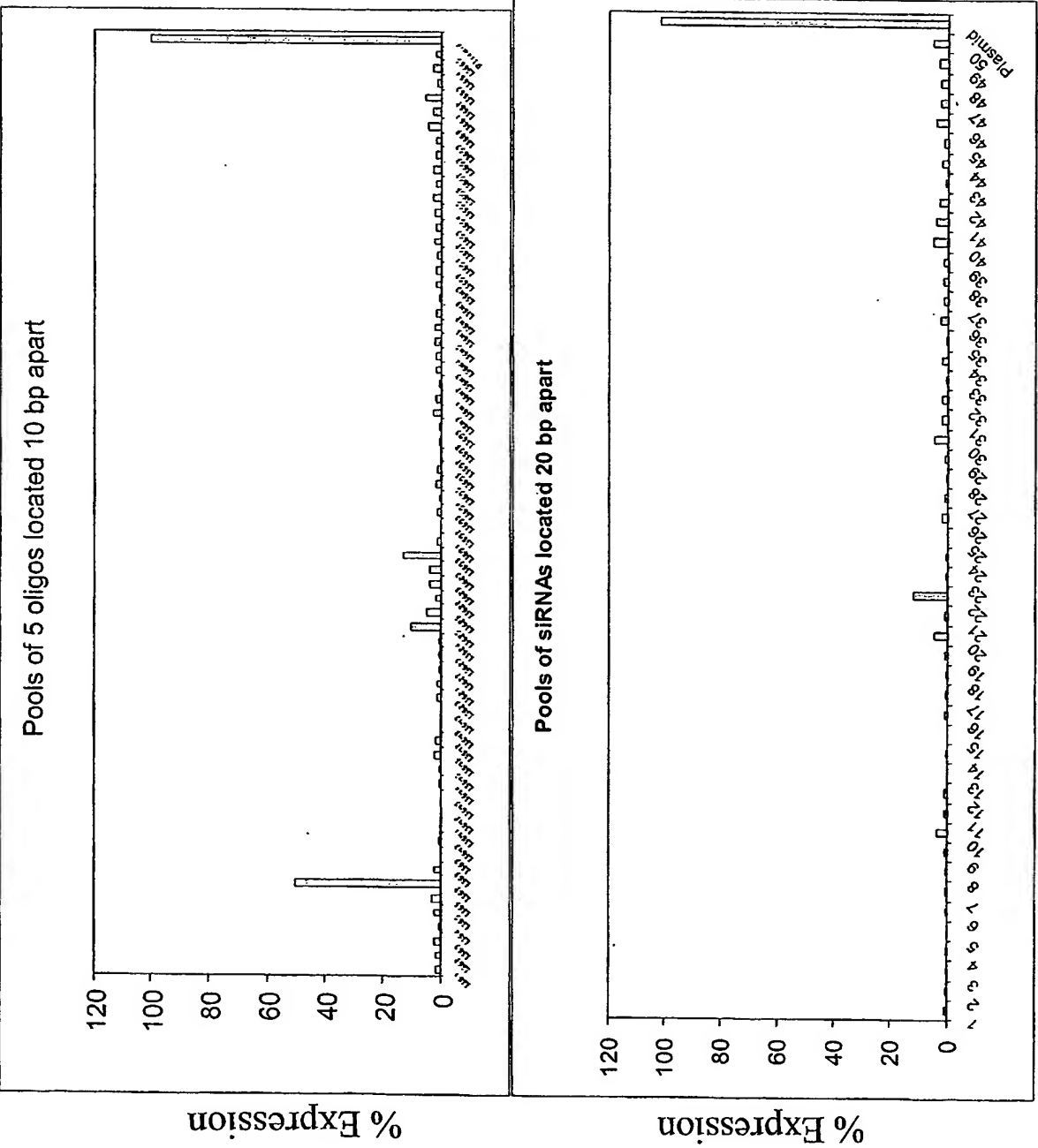


Figure 18

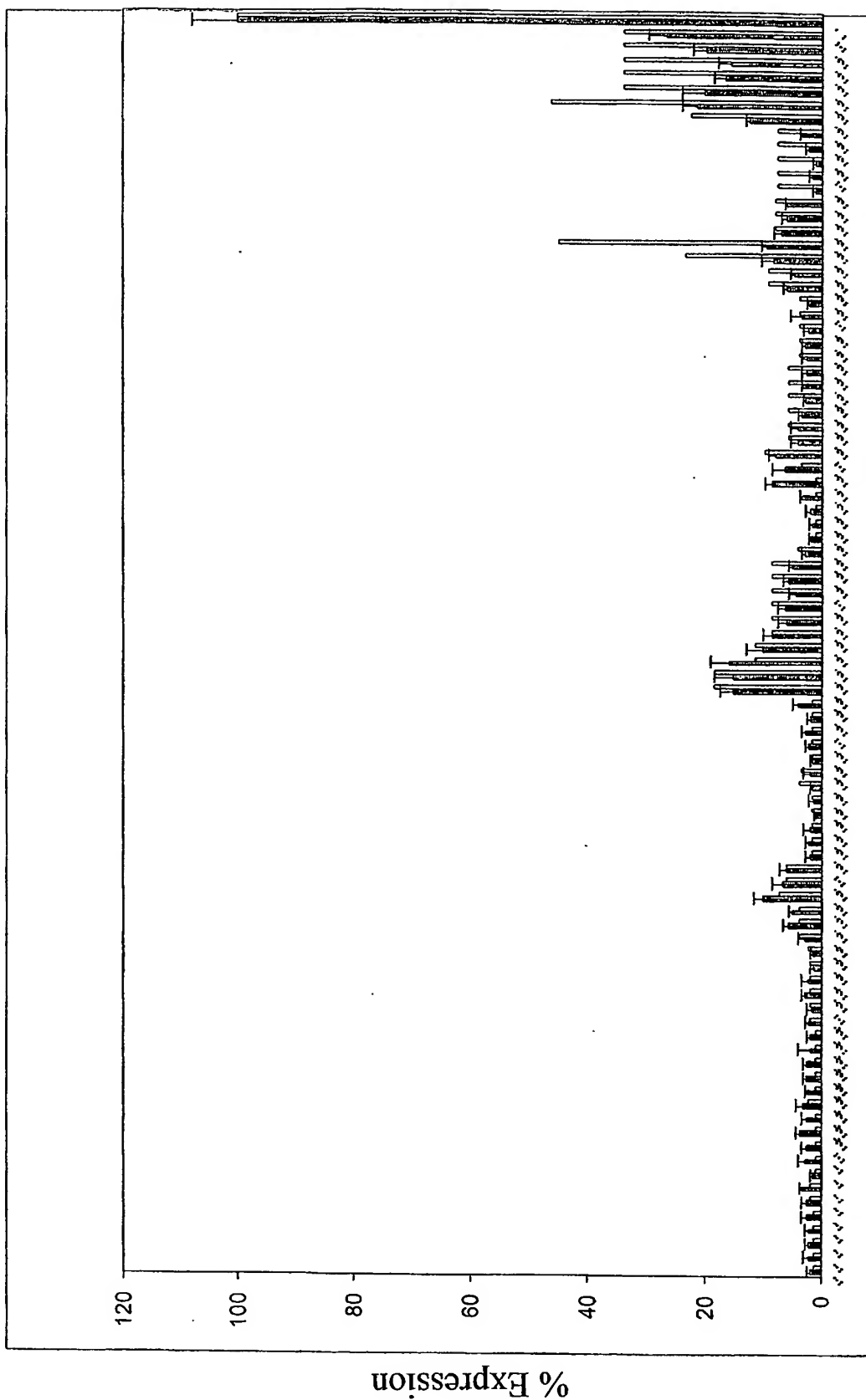


Figure 19

Combination of several semifunctional siRNAs result in a significant improvement of gene expression inhibition

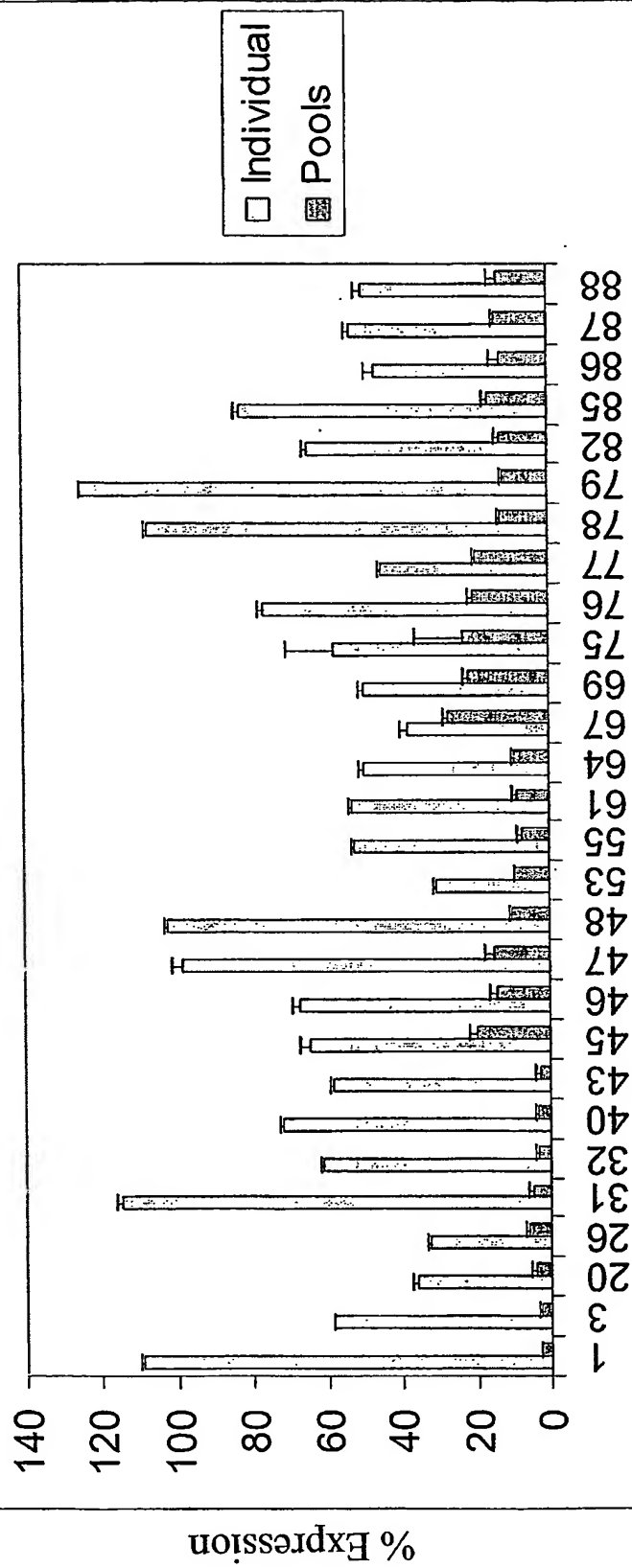


Figure 20

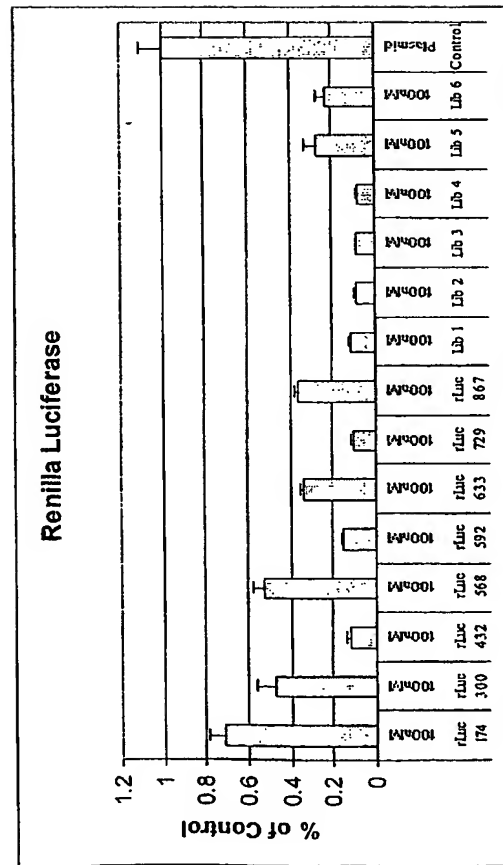
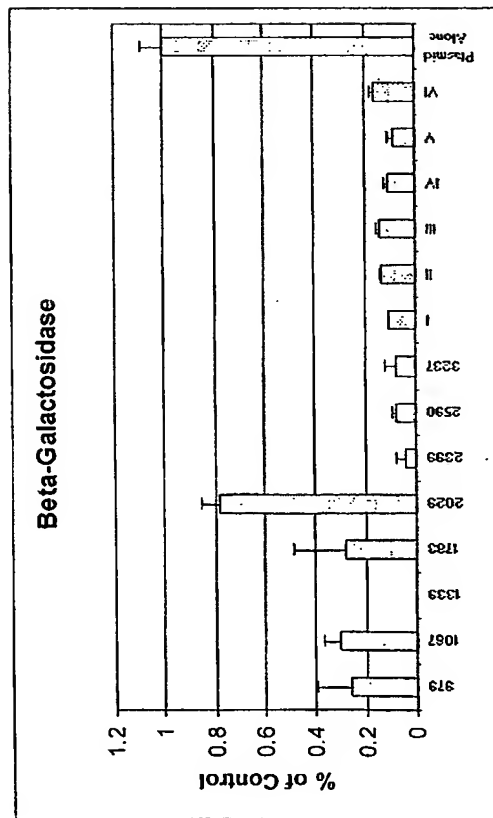
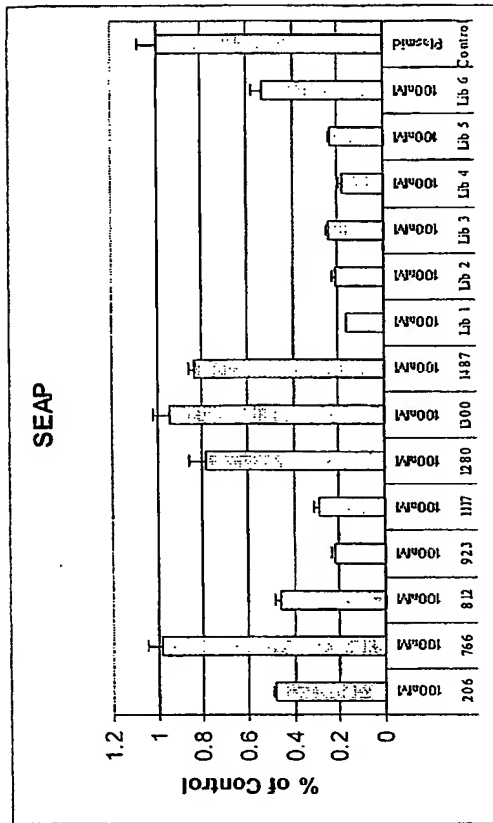


Figure 21

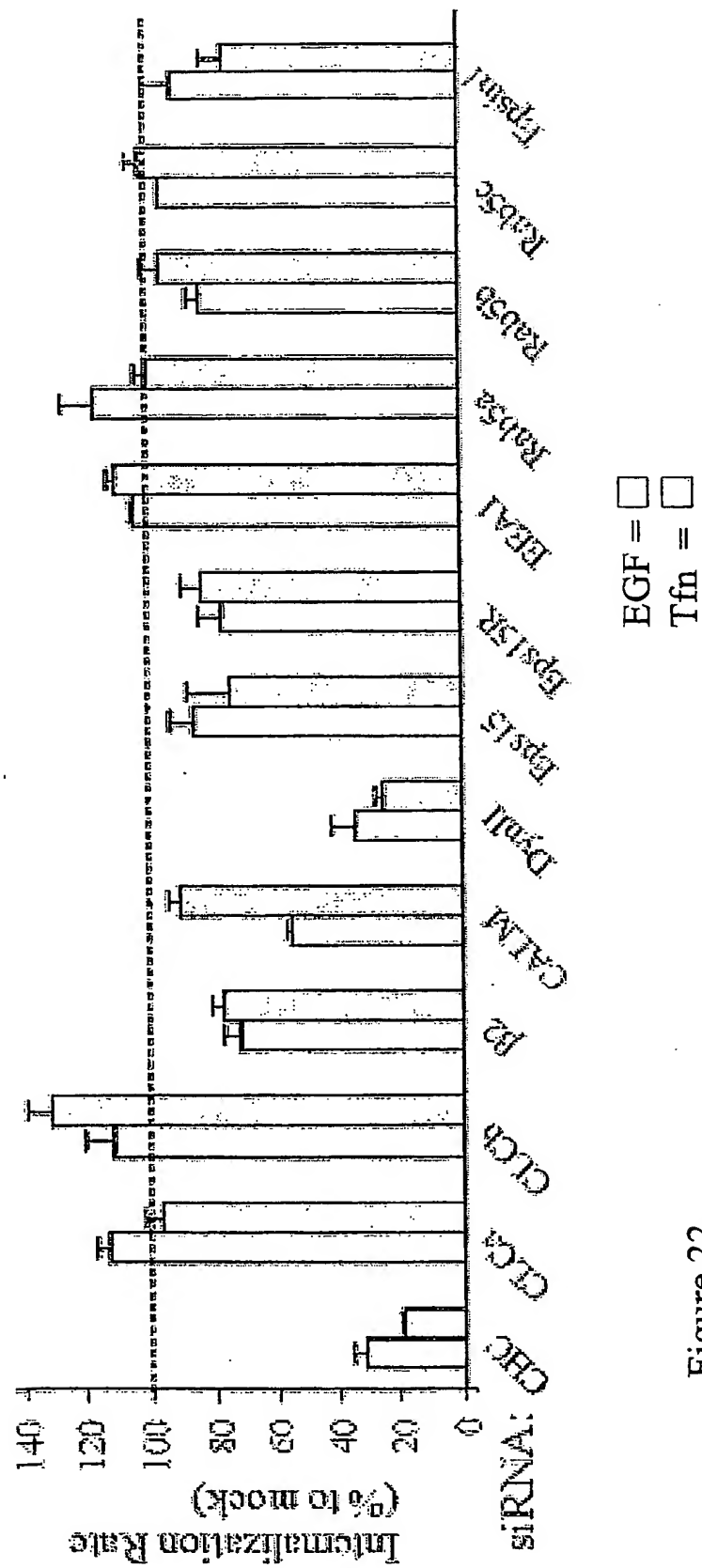


Figure 22

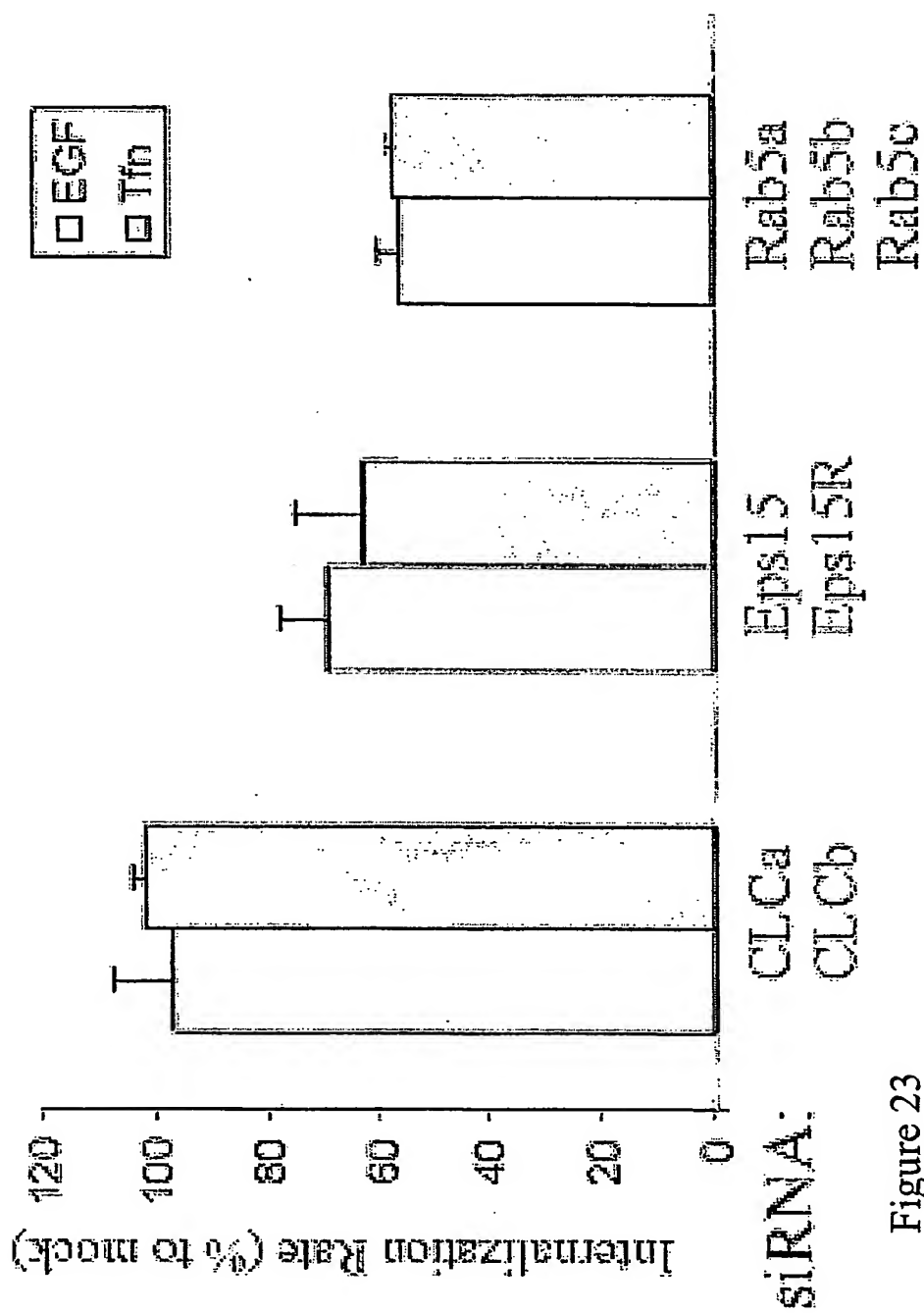


Figure 23

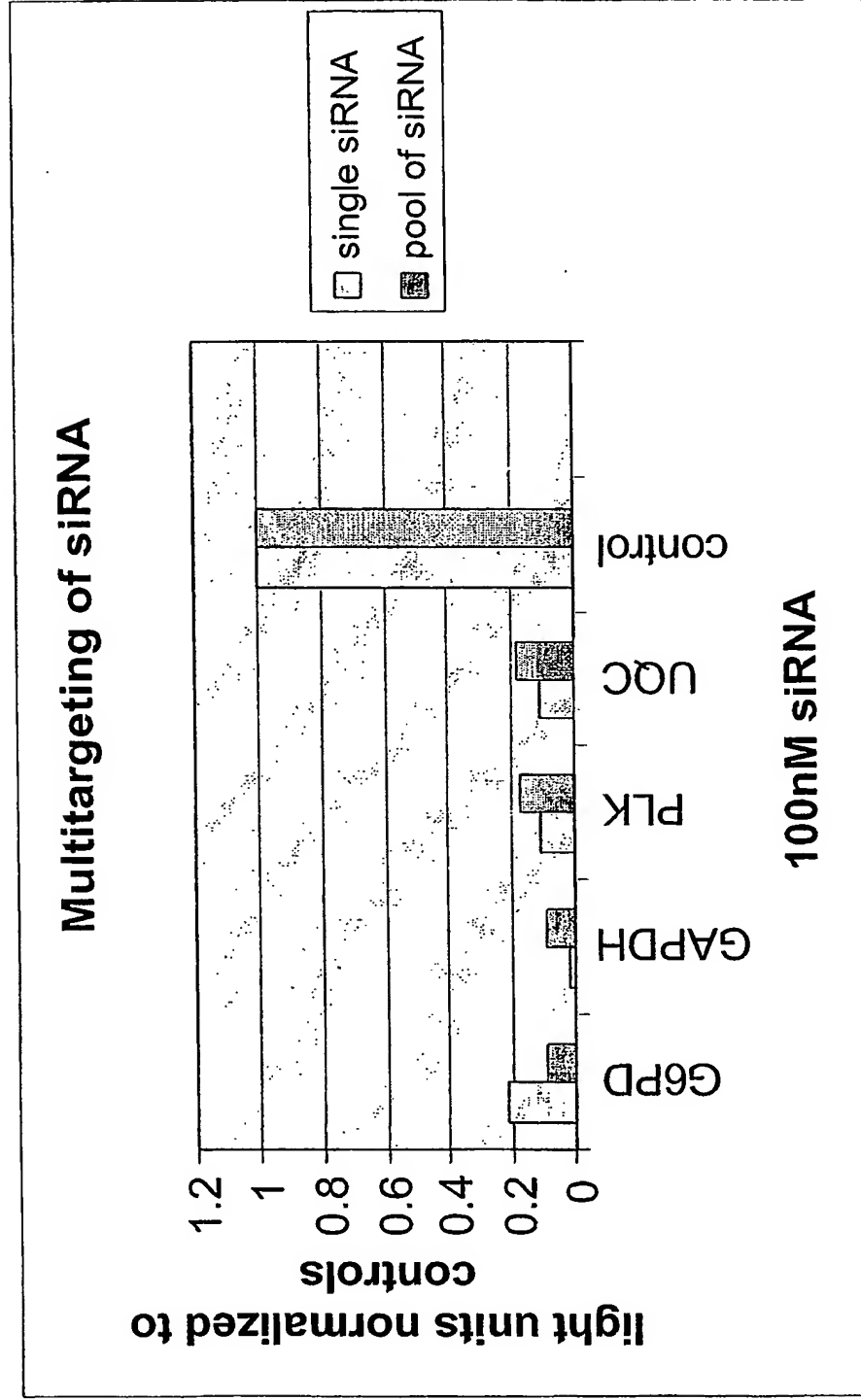
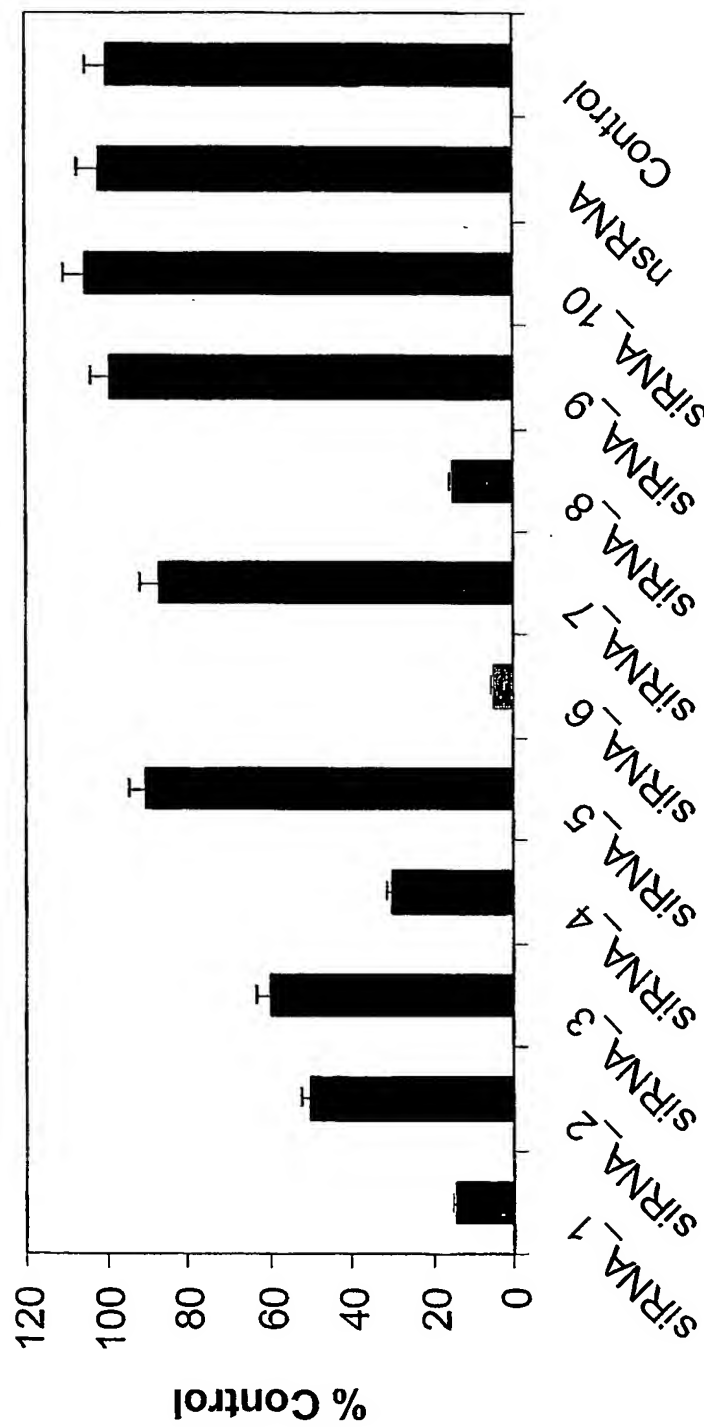


Figure 24

Figure 25

**Bcl-2 knockdown by 10 rationally designed siRNAs at
300 pM concentration**



Nucleic Acids Research, 2002, Vol. 30, No. 20 e106
© 2002 Oxford University Press

A computational framework for optimal masking in the synthesis of oligonucleotide microarrays

Simon Kasif^{*,1,2}, Zhiping Weng¹, Adnan Derti¹,
Richard Beigel^{3,4} and Charles DeLisi¹

¹ Center for Advanced Genomic Technology (CAGT),
Bioinformatics Program and Biomedical Engineering Department,
Boston University, Boston, MA, USA, ² MIT Genome Center,
Whitehead Institute, MIT, Cambridge, MA, USA and ³ Institute for
Advanced Studies, Princeton, NJ, USA and ⁴ Temple University,
Philadelphia, PA, USA

*To whom correspondence should be addressed at 44 Cummington Street, Boston University, Boston, MA 02215, USA. Tel:
+1 617 358 1845; Fax: +1 617 353 6766; Email: kasif@bu.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

Received April 3, 2002; Revised August 5, 2002; Accepted August 15, 2002

ABSTRACT

High-throughput genomic technologies are revolutionizing modern biology. In particular, DNA microarrays have become one of the most powerful tools for profiling global mRNA expression in different tissues and environmental conditions, and for detecting single nucleotide polymorphisms. The broad applicability of gene expression profiling to the

biological and medical realms has generated expanding demand for mass production of microarrays, which in turn has created considerable interest in improving the cost effectiveness of microarray fabrication techniques. We have developed the computational framework for an optimal synthesis strategy for oligonucleotide microarrays. The problem was introduced by Hubbell *et al.* Here, we formalize the problem, obtain precise bounds on its complexity and devise several computational solutions.

This Article

- ▶ [Abstract](#) FREE
- ▶ [Print PDF \(140K\)](#) FREE
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in PubMed](#)
- ▶ [Download to citation manager](#)
- ▶ [Commercial Re-use Guidelines for Open Access NAR Content](#)

PubMed

- ▶ [PubMed Citation](#)
- ▶ [Articles by Kasif, S.](#)
- ▶ [Articles by DeLisi, C.](#)

- ▲ [TOP](#)
- [ABSTRACT](#)
- ▼ [INTRODUCTION](#)
- ▼ [A COMPUTATIONAL FRAMEWORK FOR...](#)
- ▼ [HEURISTIC SEARCH SOLUTIONS](#)
- ▼ [COMPUTATIONAL ANALYSIS OF...](#)
- ▼ [SIMULATIONS WITH RANDOM OLIGOS](#)
- ▼ [CONCLUSIONS](#)
- ▼ [REFERENCES](#)

▷ INTRODUCTION

Oligonucleotide and cDNA microarrays can monitor mRNA expression levels for tens of thousands of genes simultaneously (1). While both types of arrays are applied to the elucidation of normal and pathological cellular mechanisms, the longer probes on cDNA arrays make them more susceptible to cross-hybridization, and oligo arrays are designed to reduce cross-hybridization

and improve sensitivity (2). In addition, oligo microarrays can be used to detect polymorphisms (3) and, therefore, can greatly facilitate the research and diagnosis of genetic predisposition to diseases. Several large companies, such as Affymetrix, Corning, Motorola and Samsung, have established or are establishing the capability of manufacturing microarrays in large quantities. Thus, reductions in cost or time to manufacture can have a significant impact on biotechnology and medicine.

cDNA microarrays are produced by spotting pre-made cDNA solutions onto a glass or nylon surface via physical contact or ink-jet deposition. While oligonucleotides can also be synthesized and then spotted, oligo microarrays are usually manufactured by *in situ* synthesis, primarily via photolithography (4), and more recently by ink-jet deposition (5). *In situ* synthesis involves the consecutive addition of A, C, G and T nucleotides to the appropriate spots on the microarray. An important advantage of photolithographic synthesis over ink-jet deposition is that in a single cycle of synthesis, a nucleotide can be added to all desired spots on the array. This is achieved via photodeprotection of the target spots on the array surface with UV light prior to the addition of the nucleotide. Meanwhile, non-target spots must be protected from the UV light using physical or virtual masks. The fabrication of physical masks is a laborious and costly process and one mask is needed for each cycle of synthesis for each variety of arrays. Singh-Gasson *et al.* (6) used a digital micro-mirror device to reflect light selectively onto the desired spots of an array, the 'virtual masking' strategy. Nonetheless, the deprotection step for each cycle lasts ~5 min and photolabile nucleosides are expensive. Therefore, decreasing the number of cycles required to synthesize a given set of sequences can reduce time and cost. Here, we address synthesis optimization, i.e., optimizing the order of nucleotide addition.

The simplest strategy for synthesizing a given set of sequences is to add A bases wherever appropriate as the first base, then C, G and T bases, repeating this process for the second base, and so on. Chee *et al.* (3) noted that if K is the length of the longest oligonucleotide to be synthesized, maximally $4K$ cycles are required. Hubbell *et al.* (7) observed that it would be possible to skip a synthesis cycle if a base is not needed by any oligonucleotides, or if the oligonucleotides that require the base can still be synthesized when that base is presented again later. In a parallel publication, Tolonen *et al.* (8) observe that synthesis could be accelerated, even for a large set of oligonucleotides, if the order of base addition is tailored to the oligonucleotide sequences. Consequently, oligonucleotides can vary in length by more than one base at the end of every synthesis cycle. This observation has motivated the development of the optimal base addition strategy described in this paper.

△	TOP
△	ABSTRACT
▪	INTRODUCTION
▽	A COMPUTATIONAL FRAMEWORK FOR...
▽	HEURISTIC SEARCH SOLUTIONS
▽	COMPUTATIONAL ANALYSIS OF...
▽	SIMULATIONS WITH RANDOM OLIGOS
▽	CONCLUSIONS
▽	REFERENCES

A COMPUTATIONAL FRAMEWORK FOR OPTIMAL SYNTHESIS STRATEGY

We formulate the question of devising an optimal synthesis strategy in oligo microarrays as a combinatorial state space search problem. This computational formulation provides insight into the complexity of the problem and enables a range of discrete optimization and heuristic search solutions.

▲	TOP
▲	ABSTRACT
▲	INTRODUCTION
▪	A COMPUTATIONAL FRAMEWORK FOR...
▼	HEURISTIC SEARCH SOLUTIONS
▼	COMPUTATIONAL ANALYSIS OF...
▼	SIMULATIONS WITH RANDOM OLIGOS
▼	CONCLUSIONS
▼	REFERENCES

In this paper, we assume that the input to the optimization software is a collection of N oligo sequences of arbitrary length, which have been pre-selected in a probe selection process. For simplicity, we discuss the case of uniform length K mers, but our framework is readily applicable to the more general case. An optimal synthesis strategy involves L cycles of synthesis where, in each cycle, a single and identical nucleotide is added to all unmasked oligos. The exact spatial location of each oligo on the array is not important, as long as it can be retrieved during actual synthesis. Therefore, we assume that the input to the optimization code is a list of oligos arranged in one dimension, such as shown in Figure 1A. We define a strategy for constructing K mers in L cycles as an L long vector S , consisting of elements A, C, G and T. $S[j]$ is the nucleotide added in cycle j . For instance, the vector $[A, C, A, T, G]$ corresponds to using A, C, A, T and G in cycles 1–5, respectively. We define the height of each partially constructed oligo as the number of nucleotides that have been added thus far by the synthesis strategy. We define a frontier F of a partially synthesized array to be an integer vector of size N where $F[i]$ is the height of the i th oligo thus far. For example, the frontier of the four oligos in Figure 1B is $F = [3, 1, 1, 2]$.



Figure 1. Formulation of synthesis strategy. (A) Four 3mer oligonucleotides. (B) Four partially constructed oligos, defining frontier $F = [3, 1, 1, 2]$. (C) The synthesis strategy $[A, C, G, T]$ can synthesize the array in four cycles, instead of six cycles required by the traditional approach.

[View larger version \(8K\):](#)

[\[in this window\]](#)

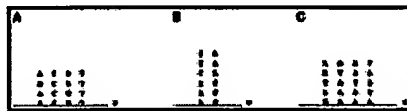
[\[in a new window\]](#)

The ‘traditional’ way to create a chip of N oligos, each of height K , is to perform $4K$ cycles of synthesis. After the addition of A, C, G and T to the appropriate spots of the array, all oligos will be one base long. We then proceed to synthesize layer two in four cycles and all oligos will be two bases long. We continue until the entire chip is synthesized in $4K$ cycles. It is easy to observe that by a slight modification of the order of base addition (8), we can expedite the above process. As a simplified

example (Fig. 1C), we can synthesize an array of $K = 3$ in four cycles using a modified synthesis strategy, compared to six cycles with the ‘traditional’ approach.

In order to introduce the optimization framework for masking, we need to measure the ‘work’ that has been accomplished after several cycles of synthesis. We therefore give two definitions of ‘frontier height’: (i) the min height of a frontier constructed after L cycles is the length of the shortest oligo [in Fig. 1B, min height (L) = 1]; (ii) the sum height of a frontier is the sum of the lengths of all oligos constructed so far [in Fig. 1B, sum height (L) = $3 + 1 + 1 + 2 = 7$].

The objective optimization criterion we desire to minimize is the number of cycles required to create a frontier of min height = K , i.e., we seek the shortest length strategy vector that is sufficient to synthesize all oligos on the chip. It is obvious that the best possible strategy for a K mer oligo chip is of length between K and $4K$. It is easy to construct an example where the shortest strategy is of length $4K$ (Fig. 2A), although genomic sequences typically do not exhibit such extremely low complexity. In general, as the number of oligos on a chip grows, the length of the optimal strategy vector is expected to grow as well.



View larger version (5K):
[\[in this window\]](#)
[\[in a new window\]](#)

Figure 2. Example arrays that challenge synthesis strategies. (A) A worst-case scenario requiring $4K$ stages for K mer synthesis. (B) The optimal solution for this chip requires 9 cycles whereas the greedy solution produces a 12-cycle strategy. However, the sum height-based greedy solution produces the optimal synthesis strategy. (C) For this example, sum height heuristics create an 8-cycle strategy: AATAATAA. The exhaustive search produces a 7-cycle strategy: ATAATAA.

► HEURISTIC SEARCH SOLUTIONS

An obvious heuristic solution to devising the optimal synthesis strategy is a greedy search. In each synthesis cycle, we consider the four different options to extend the current layout and compute the height of the resulting frontier for each option. We choose the nucleotide that maximizes the height of the frontier, which could be either min height or sum height. The min height heuristic extends the shortest oligo, while the sum height heuristic chooses the nucleotide that will add the largest number of nucleotides to the chip.

- ▲ [TOP](#)
- ▲ [ABSTRACT](#)
- ▲ [INTRODUCTION](#)
- ▲ [A COMPUTATIONAL FRAMEWORK FOR...](#)
- [HEURISTIC SEARCH SOLUTIONS](#)
- ▼ [COMPUTATIONAL ANALYSIS OF...](#)
- ▼ [SIMULATIONS WITH RANDOM OLIGOS](#)
- ▼ [CONCLUSIONS](#)
- ▼ [REFERENCES](#)

The simple examples in Figure 2B and C show that a greedy search is not guaranteed to produce an optimal solution using either the min height or the sum height heuristic. In the remainder of this section, we consider ways to improve the greedy search. In the next section, we show that an efficient

polynomial time solution is unlikely to exist for this optimization problem. We then report simulation results that describe the effectiveness of the sum height heuristics.

Look-ahead solution

A natural way to extend the greedy algorithm is to consider a look-ahead strategy. (i) Generate all possible frontiers that can be generated in L cycles. The number of strategies is 4^L . The number of frontiers might be smaller since different strategies may generate the same frontier. (ii) For each frontier, compute the height. (iii) Choose (for the first cycle) the strategy that maximizes the height after L cycles. (iv) Repeat until all oligos have been synthesized.

When $L = 4K$, this algorithm performs an exhaustive search. A rough upper bound on the running time of this algorithm is $O(4^L N)$, which makes it prohibitive for large values of L . An alternative approach would be to use a variant of best-first search such as A^* , a popular algorithm in the artificial intelligence community. A more space-efficient alternative would be to use a branch-and-bound formulation, a standard approach in discrete optimization.

Local search solution

A local search attempts to improve a given solution by a series of local perturbations until a minimum is achieved. One obvious local search approach for our problem would be to repeatedly change a selected nucleotide in the strategy vector and accept the new strategy if it results in an improvement over the previous one. Here, we implement a variant (Fig. 3) based on steepest descent.

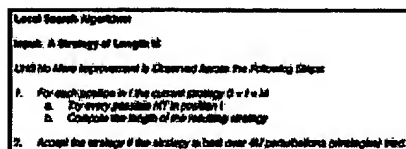


Figure 3. A local search algorithm.

View larger version (17K):

[\[in this window\]](#)

[\[in a new window\]](#)

The steepest descent algorithm considers every possible local perturbation of a single nucleotide in all positions of the strategy vector and chooses the move that results in the greatest improvement. This algorithm terminates relatively quickly since there are at most M possible improvements to be made. Each iteration (steps 1–2) requires time $O(M)$, so the total running time is $O(M^2)$.

A simple classic variant of this algorithm is to accept a new strategy with some probability even if no improvement is observed. A Gibbs sampler is a special case of this solution when the probability of acceptance is a function of the degree of improvement, and positions for possible perturbations are selected at random rather than sequentially as described above. We describe simulation results using local search below.

COMPUTATIONAL ANALYSIS OF OPTIMAL SYNTHESIS STRATEGY

In this section, we provide a set of computational reductions that allow us to obtain a precise characterization of the complexity of the optimal synthesis strategy. We see this part as the main contribution of this paper.

Multiple sequence alignment formulation

We first observe that the optimal masking problem can be reduced to a special case of multiple sequence alignment. A precise description of multiple sequence alignment can be found in Gusfield (9) and Waterman (10). In particular, the best L cycle synthesis strategy directly corresponds to the optimal multiple alignment of the N oligos, where the costs of the alignment are defined as follows: (i) replacement cost = $+\infty$; (ii) deletion cost = $+\infty$; (iii) insertion cost = $+1$. That is, the only allowed 'editing' operation is the insertion of a gap. We first demonstrate this principle with an example. For the oligo design problem in Figure 2B, we first align the two oligos CCCAAA and AAACCC. An optimal alignment is given by

CCCAAA

AAACCC

Walking across the alignment from left to right creates the following synthesis strategy: [CCCAAACCC]. Another optimal strategy is [AAACCCAAA].

The formal proof of this equivalence is not difficult. Each strategy corresponds to a multiple alignment obtained by aligning each sequence against the strategy sequence. Therefore, the shortest strategy corresponds to the shortest global alignment.

This observation enables the application of computational solutions developed in multiple sequence analysis such as dynamic programming, Gibbs sampling and iterative refinement (9,10). A common greedy solution is based on aligning each pair, then producing a strategy based on the best aligned pair and subsequently continuing to add oligos to the alignment in the best-first manner.

Shortest super-sequence formulation

The above observation is useful for obtaining insight into the problem. By reducing our problem to a special case of multiple sequence alignment, we show that multiple alignment is 'harder', which does not preclude the possibility of an efficient solution to our specific problem. Here we show that the optimal synthesis strategy problem is exactly equivalent to the problem of computing the shortest super-sequence of a collection of strings. This two-way reduction establishes our problem to be as hard as the shortest super-sequence problem, which is known to be NP hard.

▲	TOP
▲	ABSTRACT
▲	INTRODUCTION
▲	A COMPUTATIONAL FRAMEWORK FOR...
▲	HEURISTIC SEARCH SOLUTIONS
•	COMPUTATIONAL ANALYSIS OF...
▼	SIMULATIONS WITH RANDOM OLIGOS
▼	CONCLUSIONS
▼	REFERENCES

We informally define sequence X to be a super-sequence of sequence Y if every character of Y occurs in X in the same order as they occur in Y . Similarly, we define a super-sequence of a collection of sequences where the above condition has to hold for each of the sequences. For instance, AAAACCCTTTT is a super-sequence of ACT, AAACCCT and AAAATTT. Sequence X is the shortest super-sequence of a collection of sequences if and only if its length is the shortest among all super-sequences of the collection.

Since the synthesis strategy must be a super-sequence of each of the oligos, the optimal synthesis strategy vector is equivalent to the shortest super-sequence of all oligos. The shortest super-sequence problem is known to be NP hard (11) and therefore the reduction above formally establishes the optimal synthesis strategy problem to be NP hard. More explicitly, the problem of finding a masking strategy of length L ($L < 4K$), given a collection of N oligos of length K , is NP complete.

This is an important observation since it implies that the optimal synthesis strategy is unlikely to have efficient (sub-exponential) optimal solutions for a large number of oligos. It is easy, of course, to devise relatively efficient dynamic programming solutions when the number of oligos is constant (e.g. less than 10).

► SIMULATIONS WITH RANDOM OLIGOS

We have conducted a large number of simulations to estimate the performance of several heuristic approaches to devising an optimal synthesis strategy. Here we report our results with three heuristic approaches. (i) Oblivious strategy: we simply repeat synthesizing ACGTACGT... independent of the input sequences. (ii) Max sum height heuristics: we choose the nucleotide that maximizes the sum height in the next cycle (as outlined above). (iii) Randomized local search improvement: once a solution is obtained by the above two methods, we attempt to improve the solution using local search.

▲	TOP
▲	ABSTRACT
▲	INTRODUCTION
▲	A COMPUTATIONAL FRAMEWORK FOR...
▲	HEURISTIC SEARCH SOLUTIONS
▲	COMPUTATIONAL ANALYSIS OF...
▪	SIMULATIONS WITH RANDOM OLIGOS
▼	CONCLUSIONS
▼	REFERENCES

Our results comparing the oblivious and max sum height heuristics are summarized in Table 1. It is clear that for random oligos there is no significant difference in performance between max height and oblivious heuristics. It is not particularly surprising for random oligos since, roughly speaking, every layer in the chip contains an approximately equal number of nucleotides of each type (A, C, G and T). Moreover, our results for 'real' oligos appear to be consistent with this performance (data not shown). Note that one of the criteria for selecting oligos aims to prevent cross-hybridization between mRNA and multiple oligos. This puts 'selective pressure' on the design to ensure that oligos are as different as possible. As the number of oligos on the chips grows, they behave more and more like random oligos.

View this table: Table 1. Comparison of the oblivious strategy with the max sum height

[\[in this window\]](#) heuristic
[\[in a new window\]](#)

We have interpolated the expected length of a strategy for 10 000 oligos and it appears to fit the following function well:

$$f(K) = 2.5K = 4.04\sqrt{K}$$

where K is the height of the oligos.

Figure 4 shows an essentially linear fit of the data. The graph was produced by fitting the function $f(K) = 2.5K + C\sqrt{K}$, where C is the single adjustable parameter. As a result, the fit is linear. The formal derivation that proves this expectation for max sum height is implied by the analysis in Jiang and Li (11).

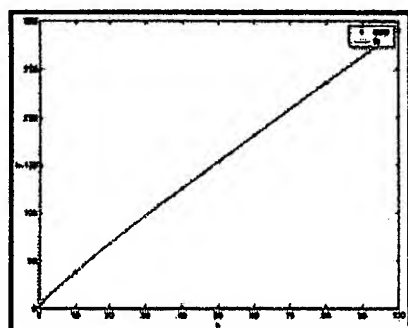


Figure 4. A linear fit of the function $f(K) = 2.5K + 4.04\sqrt{K}$ to the simulation data produced by the oblivious strategy for 1000 oligos. The scaling constant (4.04) depends on the total number of oligos.

View larger version (10K):

[\[in this window\]](#)

[\[in a new window\]](#)

Now we provide a brief motivation for the above interpolating function. When we find the shortest alignment of a single random oligo $X_1X_2...X_K$ with ACGTACGT..., X_1 aligns with the first, second, third or fourth base, X_2 aligns with the first, second, third or fourth base after X_1 , X_3 aligns with the first, second, third or fourth base after X_2 , etc. Thus the expected distance between X_i and X_{i+1} is $(1 + 2 + 3 + 4)/4 = 2.5$. For example, if $X_i = C$, then the distance between X_i and X_{i+1} is 1 if $X_{i+1} = T$, 2 if $X_{i+1} = G$, 3 if $X_{i+1} = A$ or 4 if $X_{i+1} = C$. The variance of the above possible distances is 1.25. Therefore, we expect $X_1X_2...X_K$ to require an alignment of length $2.5K$ and, by the law of large numbers, a random oligo $X_1X_2...X_K$ requires an alignment of length at most $2.5K + O(\sqrt{K})$ with high probability. In order to align N oligos, length $2.5K + O(\log N \sqrt{K})$ suffices with high probability. For a fixed number of oligos, the logarithm term is constant. Therefore, we can use the formula $2.5K + O(\sqrt{K})$.

We also used a simple local search to improve the solutions produced by the max sum height and oblivious strategies. The results are given in Table 2. We found that when the number of oligos is small (e.g. 100), the improvement was better (~3–5%). However, as the number of oligos grows (e.g. 1000), the percentage of improvement was reduced to ~1–1.5%. This is interesting, since Gibbs sampling is a close relative of local search and is a popular algorithm for multiple sequence alignment. However, in typical multiple alignments of proteins, we often align tens to hundreds of sequences. In this paper, we need to ‘align’ thousands to hundreds of thousands of oligos and it appears to have an impact on the degree of improvement obtained with this approach.

View this table: [Table 2. Results of local search improvement \(LS\) for both max sum height \[in this window\] \(MSH\) and oblivious solutions \(ACGT\) \[in a new window\]](#)

► CONCLUSIONS

In this paper, we have presented a computational formalization of the optimal synthesis strategy for oligonucleotide microarrays. We have shown that the problem is computationally intractable (NP complete). We have provided several simulation results that shed light on its practical complexity. As the number of applications of oligo microarrays increases and their

use in diagnostic medical applications becomes a common practice, we expect the design of DNA chips to become more sophisticated and efficient. Our main conclusion from both the theoretical and simulation analyses provided in this paper is that the problem of optimal masking appears to be computationally difficult. Moreover, the simplest possible solution appears to work almost as well as more sophisticated approaches that include heuristic greedy approaches and local search. It would be interesting to see if more exhaustive approaches based on best-first search, branch-and-bound or Gibbs sampling methods will generate a more dramatic improvement in performance. Naturally, these results must be confirmed in the context of practically used DNA chips.

▲ [TOP](#)
 ▲ [ABSTRACT](#)
 ▲ [INTRODUCTION](#)
 ▲ [A COMPUTATIONAL FRAMEWORK FOR...](#)
 ▲ [HEURISTIC SEARCH SOLUTIONS](#)
 ▲ [COMPUTATIONAL ANALYSIS OF...](#)
 ▲ [SIMULATIONS WITH RANDOM OLIGOS](#)
 • [CONCLUSIONS](#)
 ▼ [REFERENCES](#)

► ACKNOWLEDGEMENTS

S.K. is supported in part by NSF grants IRI-9616254 and KDI-9980088. Z.W. is supported in part by NSF grants DBI-0078194 and DBI-0078194. R.B. is supported in part by NSF grants CCR-9996021 and CCR-0019019 and a grant from the State of New Jersey. A.D. was supported by the Paul and Daisy Soros Fellowship for New Americans.



REFERENCES

▲ TOP
▲ ABSTRACT
▲ INTRODUCTION
▲ A COMPUTATIONAL FRAMEWORK FOR...
▲ HEURISTIC SEARCH SOLUTIONS
▲ COMPUTATIONAL ANALYSIS OF...
▲ SIMULATIONS WITH RANDOM OLIGOS
▲ CONCLUSIONS
▪ REFERENCES

1. Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470. [\[Abstract\]](#)
2. Lockhart,D.J., Dong,H., Byrne,M.C., Follettie,M.T., Gallo,M.V., Chee,M.S., Mittmann,M., Wang,C., Kobayashi,M., Horton,H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680. [\[ISI\]](#)[\[Medline\]](#)
3. Chee,M., Yang,R., Hubbell,E., Berno,A., Huang,X.C., Stern,D., Winkler,J., Lockhart,D.J., Morris,M.S. and Fodor,S.P. (1996) Accessing genetic information with high-density DNA arrays. *Science*, **274**, 610–614. [\[Abstract/Free Full Text\]](#)
4. Lipshutz,R.J., Fodor,S.P., Gingeras,T.R. and Lockhart,D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21** (suppl. 1), 20–24. [\[ISI\]](#)[\[Medline\]](#)
5. Blanchard,A. (1998) Synthetic DNA arrays. *Genet. Eng.*, **20**, 111–123.
6. Singh-Gasson,S., Green,R.D., Yue,Y., Nelson,C., Blattner,F., Sussman,M.R. and Cerrina,F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.*, **17**, 974–978. [\[ISI\]](#)[\[Medline\]](#)
7. Hubbell,E.A., Morris,M.S. and Winkler,J.L. (1999) Computer-aided engineering system for design of sequence arrays and lithographic masks. US Patent no. 5 856 101.
8. Tolonen,A.C., Albeanu,D.F., Corbett, J.F., Handley,H., Henson,C. and Malik,P. (2002) Optimized *in situ* construction of oligomers on an array surface. *Nucleic Acids Res.*, **30**, e107. [\[Abstract/Free Full Text\]](#)
9. Gusfield,D. (1997) *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, New York, NY.
10. Waterman,M.S. (1995) *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman and Hall, New York, NY.
11. Jiang,T. and Li,M. (1997) On the approximation of shortest common supersequences and longest common subsequences. *SIAM J. Comput.*, **24**, 1122–1139.

This Article

► [Abstract](#) **FREE**

- ▶ [Print PDF \(140K\) FREE](#)
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in PubMed](#)
- ▶ [Download to citation manager](#)
- ▶ [Commercial Re-use Guidelines for Open Access NAR Content](#)

PubMed

- ▶ [PubMed Citation](#)
- ▶ [Articles by Kasif, S.](#)
- ▶ [Articles by DeLisi, C.](#)



Nucleic Acids Research, 2000, Vol. 28, No. 21 4113-4124

© 2000 Oxford University Press

This Article

- ▶ [Abstract](#) **FREE**
- ▶ [Print PDF \(331K\)](#) **FREE**
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in ISI Web of Science](#)
- ▶ [Similar articles in PubMed](#)
- ▶ [Download to citation manager](#)
- ▶ [Cited by other online articles](#)
- ▶ [Search for citing articles in: ISI Web of Science \(34\)](#)
- ▶ [Commercial Re-use Guidelines for Open Access NAR Content](#)

PubMed

- ▶ [PubMed Citation](#)
- ▶ [Articles by Amarzguloui, M.](#)
- ▶ [Articles by Prydz, H.](#)

Secondary structure prediction and *in vitro* accessibility of mRNA as tools in the selection of target sites for ribozymes

Mohammed Amarzguioui, Gaute Brede, Eshrat Babaie, Morten Grøtli¹, Brian Sproat² and Hans Prydz^{*}

The Biotechnology Centre of Oslo, University of Oslo, Gaustadalleen 21, 0349 Oslo, Norway, ¹Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark and ²Institut für Organische Chemie, Universität Göttingen, Tammann strasse 2, 37077 Göttingen, Germany

Received July 27, 2000; Revised and Accepted September 18, 2000.

DDBJ/EMBL/GenBank accession no. [AJ272212](#).

▶ ABSTRACT

We have investigated the relative merits of two commonly used methods for target site selection for ribozymes: secondary structure prediction (MFold program) and *in vitro* accessibility assays. A total of eight methylated ribozymes with DNA arms were synthesized and analyzed in a transient co-transfection assay in HeLa cells. Residual expression levels ranging from 23 to 72% were obtained with anti-PSKH1 ribozymes compared to cells transfected with an irrelevant control ribozyme. Ribozyme efficacy depended on both ribozyme concentration and the steady state expression levels of the target mRNA. Allylated ribozymes against a subset of the target sites generally displayed poorer efficacy than their methylated counterparts. This effect appeared to be influenced by *in vivo* accessibility of the target site. Ribozymes designed on the basis of either selection method displayed a wide range of efficacies with no significant differences in the average activities of the two groups of ribozymes. While *in vitro* accessibility assays had limited predictive power, there was a significant correlation between certain features of the predicted secondary structure of the target sequence and the efficacy of the corresponding ribozyme. Specifically, ribozyme efficacy appeared to be positively correlated with the presence of short stem regions and helices of low stability within their target sequences. There were no correlations with predicted free energy or loop

- ▲ [TOP](#)
- [ABSTRACT](#)
- ▼ [INTRODUCTION](#)
- ▼ [MATERIALS AND METHODS](#)
- ▼ [RESULTS](#)
- ▼ [DISCUSSION](#)
- ▼ [REFERENCES](#)

length.

▷ INTRODUCTION

Hammerhead ribozymes are potentially powerful tools for sequence-specific inhibition of target gene expression (1). Their intrinsic cleavage activity makes them theoretically superior to traditional antisense oligodeoxynucleotides (ODNs) in terms of inhibitory capacity. Recent advances (1–3) have extended the range of targets

so far that virtually any limited stretch of RNA is now likely to contain a useful target. However, other problems, including methods of delivery and target site selection, remain to be solved. The latter, in particular, appears to be a critical step in the design of antisense or ribozyme molecules for suppression of target gene expression, as there may exist only a few sites within any mRNA that are accessible to hybridization (4,5). We have compared two commonly employed methods for the rational selection of target sites, secondary structure prediction and *in vitro* accessibility assays. As target we have used the mRNA for a novel human protein kinase, PSKH1 (6).

△	TOP
△	ABSTRACT
◦	INTRODUCTION
▽	MATERIALS AND METHODS
▽	RESULTS
▽	DISCUSSION
▽	REFERENCES

RNA secondary structure prediction algorithms based mainly on energy minimization have intrinsic limitations, although subject to continuous improvement (7,8). An established theory for selection of target sites based on predicted structure is also required. The possible modulation of RNA structures by protein binding *in vivo* (9,10) cannot yet be modeled and has so far limited their use in designing antisense ODNs or ribozymes. A positive correlation between the inhibitory effect of an antisense RNA and low local folding potential has been noted (high ΔG) (11). Together with more recent data (12–14), this suggests that extended single-stranded or unstructured regions may be the best targets for antisense ODNs. However, although ribozymes targeting predicted loop regions have proved efficacious in some cases (15,16), there are also examples of failures (17). A recent systematic analysis of the hybridization of tRNA-Phe to a set of complementary ODNs determined that all high-yield heteroduplexes involved RNA sequences forming both double-stranded stems and single-stranded regions, and that bases of the latter regions were often stacked onto the stems (18). This suggests the requirement for at least some degree of helical conformation in the secondary structures of favorable targets.

The accessibility of different stretches of the mRNA for hybridization with short antisense ODNs may be determined *in vitro* by RNase H-mediated cleavage of the RNA at regions where the ODNs hybridize to the target transcript (14,19–22). Screening of a large set of ODNs targeting potential ribozyme cleavage sites has been employed to select the most promising sites for ribozyme targeting (19). Accessibility assays with specific ODNs have also been performed in cell extracts on endogenous transcripts in order to better approximate the *in vivo* situation (14,22). The most accessible sites within any target RNA may be selected by performing the RNase H-assay with a randomized set of ODNs (20,21) or with a target-specific set of cDNA fragments prepared by partial DNase I-digestion (23). ODN libraries in previous studies (20,21) have been more or less completely randomized. We wanted to identify specific sites amenable to cleavage with standard hammerhead ribozymes, and to restrict the analysis to those triplets that are cleaved most efficiently. Since the hammerhead ribozyme seems to

prefer a purine in the first position of the cleavage triplet and a C or an A in the third position (24), we decided to concentrate on the triplets GUC, GUA, AUC and AUA in our analysis. Four ODN libraries, each specific for one of these cleavage triplets, were synthesized and used to screen *in vitro* transcribed PSKH1 RNA for accessible sites.

Based on a study of the most active ribozymes expressed *in vivo* from a randomized library (25), the ribozymes used in the present study were designed to have symmetric 8+8 nt arms (Fig. 1) (26). Chemically synthesized ribozymes of similar arm lengths have subsequently been successfully employed (19,27–29). The stem II structure (Fig. 1) was chosen to be 2bp long (30). The activity of such truncated ribozymes in cell culture has been demonstrated (27,29). For increased nuclease stability of the ribozymes, we have retained unmodified ribonucleotides in only five catalytically important positions (31,32). Deoxyribonucleotides were used in the flanking arms (33,34) and 2'-*O*-alkylated ribonucleotides in the core and stem-loop II (31,32,35) (Fig. 1). Nuclease stability of ribozymes was increased further by an inverted thymidine at the 3'-end (28,29,32) and by a hexanol moiety at the extreme 5'-end (36). Ribozymes used *in vivo* frequently include short stretches of phosphorothioate linkages at the 5'-end (28,29), 3'-end (34) or both (27) for stabilization against exonucleases. We included two phosphorothioates at the 5'-end and one at the 3'-end. DNA nucleotides in the arms have been reported to increase catalytic efficiencies of ribozymes *in vitro*, most likely due to increased product dissociation rates (37–39). DNA-armed ribozymes may recruit RNase H activity upon hybridization with the target RNA and thus enhance their apparent activity (40). We are here primarily interested in the accessibility of cleavage sites. Anything that reduces the importance of intrinsic ribozyme cleavage activity and increases the importance of target-specific inhibition of expression is desirable.

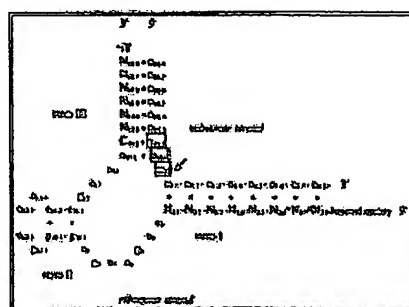


Figure 1. Schematic of the hammerhead ribozyme hybridizing to its target mRNA. The arrow indicates the position of cleavage in the mRNA. Numbering is according to the nomenclature of Hertel *et al.* (26). Unmodified ribonucleotides are in bold lower case, 2'-*O*-alkylated ribonucleotides in plain lower case, and deoxynucleotides in upper case. Phosphorothioates are indicated by asterisk, while iT denotes an inverted 3'–3' thymidine. Bases of the GUM target triplet, where M is C or A, are boxed.

View larger version (14K):
[\[in this window\]](#)
[\[in a new window\]](#)

Cellular delivery of ribozymes is commonly accomplished with various cationic liposome formulations (16,19,27–29). We have used the cationic liposome reagent lipofectamine to co-transfect HeLa cells with a mixture of ribozyme, a firefly luciferase reporter gene construct containing the complete coding cDNA of the target PSKH1, and a *Renilla* luciferase-encoding plasmid serving as an internal transfection control. We have succeeded in constructing a ribozyme against PSKH1 mRNA which

reduced the activity of the corresponding reporter gene to 20–25% in a concentration dependent manner. A correlation is observed with certain features of the predicted secondary structure of the target mRNA.

▷ MATERIALS AND METHODS

All restriction enzymes were from New England Biolabs.

Synthesis and purification of hammerhead ribozymes

Automated RNA and DNA synthesis was carried out on an Applied Biosystems model 394 DNA/RNA synthesizer. 2'-O-Alkyl ribozymes containing five unmodified purine ribonucleotides were synthesized by solid phase β -cyanoethyl phosphoramidite chemistry (41), using the 2'-O-*tert*-butyldimethylsilyl protection strategy for the ribonucleotides (42,43). Syntheses were performed on controlled pore glass bearing an inverted thymidine linkage (Glen Research). A lipophilic capture tag, 1-[Diisopropyl(DL- α -tocopheryloxy)silyl]oxy-6-(2-cyanoethyl *N,N*-diisopropylamino-phosphinoxy) hexane, was added at the 5'-end of the oligomer as described (36). Cleavage from the support and release of all base labile protecting groups (44,45), reverse phase HPLC purification (Pharmacia Source 5RPC 10/10 column, using a flow-rate of 1 ml/min, or a μ Bondapac C-18 column), lyophilization, desilylation (46), butanol-precipitation and counter-ion exchange with NaClO₄ were performed essentially as described (36). Ribozymes, retaining a 5' hexanol-linker, were desalted (NAP-10 columns, Pharmacia) and quantified by UV spectroscopy. Molar extinction coefficients were calculated based on the nearest-neighbor method (Biopolymer Calculator at <http://paris.chem.yale.edu>). Exact molecular weights were calculated. Ribozymes were controlled by denaturing 15% polyacrylamide gel electrophoresis prior to their application in cell culture experiments.

Sequence and modification of hammerhead ribozymes

PSKH1-specific 2'-O-methylated ribozymes (MRz) were designed targeting a total of eight sites. Allylated ribozymes (ARz) targeting a subset of these sites were also synthesized in order to evaluate the relative effects of the two types of modification on the efficacy of ribozymes. For both types of ribozymes, controls having the same chemical composition and length of hybridizing arms were designed targeting an irrelevant mRNA (human tissue factor). Ribozymes are named after their type of modification and cleavage position. Thus MRz-519 is a methylated ribozyme cleaving after *guc519*, while ARz-519 is the corresponding allylated ribozyme. The respective control ribozymes are MRz-TF and ARz-TF. The unique sequences (flanking arms) of the ribozymes were as follows (the conserved sequence of the core is indicated by <core> for all but the first ribozyme sequence): MRz-TF/ARz-TF, A*A*T-C-T-C-C-T-c-u-g-a-u-g-a-g-g-u-u-a-c-c-g-a-a-a-C-T-T-A-G-T*G-iT; MRz-118, T*C*G-G-G-A-A-G- <core> -a-C-C-T-T-G-C*T-iT; MRz-238, C*C*G-G-C-T-T-T <core> -a-C-A-G-G-G-C*C-iT; MRz-287, A*G*T-C-G-G-G-G- <core> -a-C-C-G-G-G-G*C-iT; MRz-465/ARz-465, T*G*A-T-G-G-C-A- <core> -a-C-G-G-C-T-G*C-iT; MRz-519/ARz-519, G*C*A-G-C-T-C-C- <core> -a-C-T-C-A-C-A*C-iT; MRz-539/ARz-539, A*C*G-C-A-C-C-C- <core> -a-C-G-C-A-G-C*A-iT; MRz-548, G*T*T-T-G-G-C-A-T- <core> -a-C-G-C-A-C-C*C-iT; MRz-712, A*G*A-T-A-C-C-G- <core> -a-C-G-C-C-A-T*C-iT. Modified and unmodified ribonucleotides are in plain and bold lower case, respectively, while

△	TOP
△	ABSTRACT
△	INTRODUCTION
▫	MATERIALS AND METHODS
▽	RESULTS
▽	DISCUSSION
▽	REFERENCES

deoxyribonucleotides are in upper case. An asterisk denotes a phosphorothioate linkage, while iT indicates an inverted 3'-3' thymidine (Fig. 1).

Specific and semi-randomized DNA ODNs

Semi-randomized 13mer antisense ODNs specific for each of the target triplets GUC, GUA, AUC and AUA were synthesized and HPLC-purified. The ODNs were designed to mimic the hybridizing arms of symmetrically armed (6+6 nt) hammerhead ribozymes and had the following sequences (N denotes a randomized position): GUC-specific library, (N)₆GAC(N)₄; GUA-specific library, (N)₆TAC(N)₄; AUC-specific library, (N)₆GAT(N)₄; AUA-specific library, (N)₆TAT(N)₄. Specific 13mer antisense ODNs targeting the selected ribozyme cleavage sites were also synthesized. These ODNs are identified by the cleavage position of their corresponding ribozymes, and have the following sequences: O-118, GGGAAGGACCTTG; O-238, GGCTTTGACAGGG; O-287, TCGGGGGACCGGG; O-465, ATGGCATACGGCT; O-519, AGCTCCGACTCAC; O-539, GCACCCGACGCAG; O-548, TGGCATGACGCAC; O-712, ATACCGGACGCCA.

Plasmids

The cDNA of the coding sequences of PSKH1 was previously cloned in our lab and inserted into the expression vector pcDNA3 (Invitrogen) downstream of a T7 promoter, producing the plasmid pcDNA3-PSK. A *Renilla* luciferase expression plasmid, pEF1-Rluc, constructed by inserting the EF1 α promoter into the multiple cloning site of the pRL-null vector (Promega), was used as an internal control in co-transfection experiments. This plasmid was a kind gift from Professor A.-B. Kolstø.

Preparation of PSKH1-luciferase fusion constructs

The full-length coding sequence of PSKH1 cDNA was cloned in-frame with firefly luciferase cDNA into the *Bgl*III/*Nco*I sites of the pGL3 Enhancer expression vector (Promega) as a *Bam*HI-*Nco*I PCR fragment, producing the plasmid pPSK-Luc. The PSK-Luc fusion was cloned into the tetracycline response plasmid pTRE (Clontech) in two steps. A *Kpn*I-*Bam*HI fragment of pPSK-Luc was first transferred to the EGFP-1 vector (Clontech). Subsequently, an *Eco*RI-*Bam*HI fragment from this clone was excised and cloned into the same sites of the pTRE vector, giving pTRE-PSK-Luc. Finally, PSK-Luc cDNA was also cloned into the pcDNA3 expression vector as an *Eco*RI-*Xba*I fragment for higher-level expression. This expression plasmid was termed pcDNA3-PSK-Luc.

In vitro transcription of PSKH1 RNA

For *in vitro* transcription of PSKH1 RNA with T7 RNA polymerase, the pcDNA3-PSK plasmid was linearized internally with *Eco*NI, producing a 1.03 kb transcript containing 47 nt of 5' vector-derived sequence, or downstream with *Xba*I (producing a 1.37 kb transcript). Protein and salt were removed from restriction reactions using JetQuick columns (Genomed). Purified DNA was eluted in DEPC-treated water (DEPC-H₂O), precipitated with ethanol, and resuspended in DEPC-H₂O. Run-off transcription of uniformly ³²P-labeled RNA was performed in 50 μ l reactions containing 2 μ g template DNA, 0.5 mM each of GTP, ATP, CTP and UTP (Boehringer Mannheim), 50 U RNasin (Promega), 1–2 μ l 10 μ Ci/ μ l [α -³²P]rATP or rUTP (Amersham) and 50 U T7 RNA polymerase (New England Biolabs) in RNA polymerase buffer (40 mM Tris-HCl, 6 mM MgCl₂, 2 mM spermidine, 10 mM DTT, pH 7.9).

Reactions were incubated for 2 h at 37°C. RNA was then desalted by centrifugation through RNase-free G50 Sephadex QuickSpin Columns (Boehringer Mannheim) and deproteinated by phenol extraction. RNA was then precipitated with isopropanol and resuspended in RNase-free water. Alternatively, RNA was purified by LiCl-precipitation (2.5 M final concentration) following transcription. Yield and concentration of RNA were calculated from the percentage of incorporated radioactivity.

Antisense ODN and RNase H-mediated *in vitro* cleavage of PSKH1 RNA

RNase H-mediated cleavage assays with antisense ODN libraries were performed in 10 µl reactions containing 40 µM ODN library, 50 nM PSKH1 mRNA and 0.25 U RNase H (Promega) in a buffer containing 40 mM Tris-HCl pH 7.5, 10 mM MgCl₂ and 10 mM DTT. A stock dilution of RNA in buffer was heated to 95°C for 60 s and then preincubated at 37°C for 15 min before adding enzyme. An 8.0 µl mixture of RNA, buffer and enzyme was then mixed with 2.0 µl 200 µM ODN library, yielding the indicated final concentrations of components. Reactions were incubated for 30 min at 37°C, and quenched on ice with 10 µl of denaturing loading buffer (8 M urea, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were heated to 95°C for 2 min prior to loading onto a 4% denaturing polyacrylamide gel. RNA size markers were prepared by run-off *in vitro* transcription of differently linearized (*Nco*I, *Xmn*I, *Bsa*A1, *Nae*I, *Bst*UI, *Fok*I and *Dde*I) pcDNA3 templates. Samples and standards were analyzed by electrophoresis for 3–4 h at 50 W at room temperature on 40 cm sequencing gels. The gels were then transferred to Whatman 3MM paper, wrapped in plastic foil and exposed overnight in a Phosphor Screen (Molecular Dynamics) prior to analysis on a radioanalytical scanner (Storm 860, Molecular Dynamics). Fragment sizes were calculated from their measured mobilities and those of the size markers. Reactions with specific antisense ODNs (200 nM, 4-fold molar excess) were performed essentially as for the ODN libraries.

Cultured cells

HeLa cells (from ATCC) were maintained in Dulbecco's Minimal Essential Medium supplemented with 2 mM glutamine and 10% fetal calf serum (all reagents from Gibco BRL). Upon thawing, the cells were grown to near confluency for 2–3 days and passaged at least once before they were used for experiments. Cells were routinely passaged every 3–4 days.

Transient co-transfections

Cells were plated in 12-well (Costar) or 24-well (Sarstedt) plates at 30–40% confluency and transfected at an estimated 60–80% confluency the following day. Transfections were performed with 100 nM ribozyme (1.23–1.30 µg/ml), 0.40 µg/ml reporter construct (pTRE-PSK-Luc or pcDNA3-PSK-Luc) and 8 ng/ml internal control plasmid (pEF1-Rluc). Nucleic acids were complexed with lipofectamine (Gibco BRL) (a final concentration of 8.2–8.5 µg/ml lipid) at a constant 1:5 w/w ratio, following optimization with the transfection agent. Complexes were formed by mixing equal volumes of medium-diluted nucleic acids (ribozyme + DNA) and lipofectamine and incubating at room temperature for 30 min. The complexes were subsequently diluted to the final volume with serum-free medium and added to pre-washed cells (250 and 500 µl for 16 and 22 mm wells, respectively) for 5 h. The complexes were subsequently replaced with full medium.

Uptake of FITC-labeled ribozymes

Uptake of FITC-labeled ribozymes under the standard transfection conditions was determined by fluorescence spectroscopy. Lipofectamine complexes were diluted with medium and chilled on ice prior to addition to cells (in 12-well plates) preincubated (for at least 30 min) either at 37°C or on ice. Following transfection for 5 h, cells were harvested by washing three times with ice cold phosphate-buffered saline (PBS) and lysed with 1.5 ml PBS-TDS (1% Triton, 0.5% deoxycholate and 1% SDS in PBS) for at least 15 min at room temperature under vigorous shaking to achieve complete lysis. Parallels were combined and fluorescence recorded on a spectrometer (LS-5, Perkin Elmer) at excitation and emission wavelengths of 492 and 522 nm, respectively.

Luciferase activity assays

Cells were harvested 24 h after initiation of transfection and washed twice in cold PBS prior to lysis for 30 min on ice with a passive lysis buffer supplied with the Dual-Luciferase® Reporter Assay System kit (Promega). Following brief centrifugation to pellet cell debris, luciferase assays were performed in white non-transparent 96-well plates (Nunc) using a plate-reading luminometer (MicroLumat Plus, EG&G Berthold) equipped with two injectors. Dual-luciferase assays were performed on 25 µl lysate supernatant. Firefly and *Renilla* luciferase activities were recorded following the respective injections of 100 µl LAR II and 100 µl Stop&Glow reagents of the assay system according to the manufacturer's instructions. The instrumental background levels of luminescence were recorded for empty wells.

▷ RESULTS

Selection of ribozyme target sites by MFold secondary structure prediction

We have used the MFold program (version 3.0 at

<http://mfold2.wustl.edu>) (7,8) to predict the secondary structure of PSKH1 mRNA as a means of selecting GUC sites for ribozyme targeting. The MFold program generated many suboptimal secondary

structures that differed minimally in energy from the optimal folding. Although these secondary structures displayed varying degrees of folding differences, certain substructures could be identified that demonstrated a considerable degree of conservation. High conservation of these substructures among alternate suboptimal structures may increase their likelihood of being 'true' structures (i.e. contained in the actual secondary structure of the mRNA *in vivo*). Therefore, in selecting GUC sites for hammerhead ribozyme targeting, we have screened 10 optimal and suboptimal secondary structures of the first 1.0 kb of the PSKH1 transcript, as transcribed from pTRE-PSK-Luc, for recurring substructures. All selected target sites were located within identical substructures in at least six out of the 10 most energetically stable structures, including the optimal structure (Table 1). For some target sites, several of the suboptimal structures diverged only slightly from the consensus in ways that did not significantly change the basic characteristics of the substructure. Due to the lack of a consensus on what kind of structures make the best targets, we have chosen GUC targets (guc238, guc287, guc519 and guc712) that are predicted to be located within very different substructures (Fig. 2). The guc238 cleavage triplet is located within a large internally looped stem in which the 5' arm (stem I) of the corresponding ribozyme is complementary to the nucleotides of the loop. The guc287 target sequence, on the other hand, is

△	<u>TOP</u>
△	<u>ABSTRACT</u>
△	<u>INTRODUCTION</u>
△	<u>MATERIALS AND METHODS</u>
▫	<u>RESULTS</u>
▽	<u>DISCUSSION</u>
▽	<u>REFERENCES</u>

predicted to fold into a short stem-loop structure. The *guc519* target site is located at the end of what is essentially a very long basepaired region that is interrupted by occasional short bulges. A ribozyme was also designed targeting *guc712*, located at a particularly stable hairpin structure at the end of a very long, and presumably very stable, stem. This target site was included due to the high conservation of its secondary structure among suboptimal structures of the target RNA (identical in 10 out of 10 and 19 out of the 20 energetically most favorable structures). In addition to the above target sites, the first GUC site in the RNA (*guc118*) was selected due to its close proximity to the translation initiation region (only ~20 nt downstream of the initiation codon). This region has often been targeted (12,33,47,48) because the RNA in this region might be relatively open due to the need for binding of the components of the translation machinery. The predicted secondary structure for the *guc118* target consists of a stem with two internal loops (Fig. 2).

View this table: [Table 1. Features of the predicted secondary structure of selected ribozyme target sites](#)
[\[in this window\]](#) [\[in a new window\]](#)

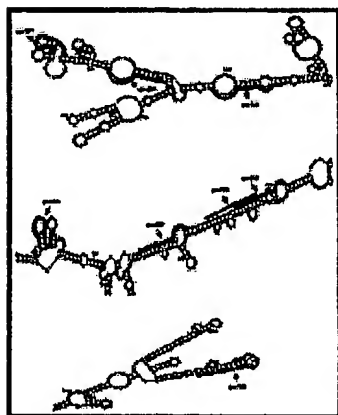


Figure 2. Excerpts of the energetically optimal secondary structure of the first 1000 bases of the pTRE-PSK-Luc transcript, as predicted by MFold version 3.0 (7,8). Target sequences of the ribozymes are highlighted. Arrows indicate position of cleavage.

View larger version (28K):
[\[in this window\]](#)
[\[in a new window\]](#)

***In vitro* accessibility assays with semi-randomized ODN libraries and specific ODNs**

Target triplet-specific, semi-randomized ODN libraries were used to screen *in vitro* transcribed PSKH1 RNA for hybridization-accessible GUC, GUA, AUC and AUA sites. Accessible sites were identified by RNase H-mediated cleavage of the RNA at the RNA-DNA hybrids generated by hybridization of antisense ODNs to target RNA (19–21). Screening of a 1.03 kb PSKH1 transcript from the pcDNA3-PSK plasmid was performed separately with each of the four ODN libraries (Fig. 3A). Since the transcript was uniformly labeled, cleavage at any position produced two cleavage products. This complicated the interpretation of the data since each pair of cleavage fragments suggested two possible sites of cleavage. Screening reactions were subsequently performed on a longer RNA transcript (1.37 kb) that differed from the first transcript in its 3'-end. This facilitated the interpretation of cleavage

data, as 5' cleavage fragments were identical for the two RNAs. Screening of two differently sized transcripts increased confidence in the results. Any accessible sites that are not identified in the longer transcript are likely to be due to structural features of the specific transcript that may not be represented in the full-length mRNA. Such sites are more likely to be encountered at the ends of the transcripts since the local folding of these regions will be most strongly affected by the lack or presence of additional sequences. Consequently, any cleavage sites originating within the first or last 200 nt of the transcript were ignored in the selection of accessible sites by library screening, as were cleavage sites that could not be detected in the longer transcript.

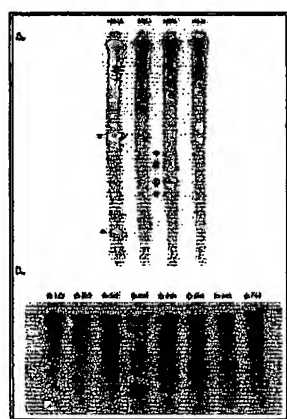


Figure 3. *In vitro* accessibility assays. (A) Separate screening of a 1.03 kb PSKH1 *in vitro* transcript with triplet-specific ODN libraries. Reactions were performed with 50 nM RNA and 40 μ M ODN library as detailed in Materials and Methods. Arrows indicate pairs of fragments resulting from cleavage at subsequently selected ribozyme target sites. (B) Cleavage of PSKH1 RNA (50 nM) with molar excess (200 nM) of specific ODNs against ribozyme target sites selected by oligo library screening (O-465, O-519, O-539, O-548) and structural considerations.

[View larger version](#)
(63K):
[\[in this window\]](#)
[\[in a new window\]](#)

The screening assays demonstrated several reproducible cleavage sites that obeyed the above restrictions. Most cleavage sites were identified with the GUA and GUC libraries, while no useful sites were identified with the AUA library (Fig. 3A). Screening of PSKH1 RNA with the GUA library produced two very prominent cleavage fragments that were not produced by any of the other libraries, demonstrating their GUA triplet specificity. The sizes of these two fragments were estimated at ~615 and 405 nt. Their combined estimated size of 1020 nt was in good agreement with the transcript length (which was 1030 nt). The shorter fragment was shown to be the 5' cleavage fragment, as it was also present when screening the longer RNA transcript (data not shown). The size of this fragment was consistent with cleavage near a site corresponding to position 460 in the PSKH1 cDNA. Examination of the sequence indicated a suitable GUA triplet at position 463–465 as the only probable site of cleavage. Screening with the GUC library produced six closely spaced fragments with estimated sizes ranging from 470 to 560 nt which fitted nicely into three pairs with combined sizes of the expected length (1020–1030 nt). By a similar analysis as for the GUA library cleavage fragments, the putative cleavage sites were identified as *guc519*, *guc539* and *guc548*, respectively. One of these sites, *guc519*, had previously also been selected on the basis of MFold secondary structure predictions (see above) prior to the antisense library screening. The AUC library screening also produced several fragment pairs, but these were either weaker than their GUA and GUC counterparts, or could not be unambiguously

assigned to a specific triplet. The screening assays were performed with a ratio of ODN (40 μ M) to target RNA (50 nM) of 800:1, which given the complexity ($4^{10} = 1 \times 10^6$) of the libraries corresponds to a ratio of specific ODN (40 pmol) to RNA of roughly 1:1250. This might seem too low a ratio to reasonably account for the degree of cleavage observed. However, the observed cleavage would correspond to a turnover of only ~ 100 molecules/h. Furthermore, the concentration of ODNs that can productively hybridize to a given target site may be significantly higher as mismatches corresponding to the ends of the ODNs are not expected to severely impair hybridization efficiency. Non-triplet-specific cleavage events (involving ODNs with mismatched or wobble-paired triplets, as well as partial hybridization of only the random-armed portions) may also conceivably occur, although they are less likely to account for the major cleavage events. Non-specific cleavage events would also be expected to result from screening with more than just one library. Examples of such cleavage events were observed, but did not include the selected cleavage sites. Nonetheless, to ensure that putative ribozyme target sites were properly identified, the provisional identifications were subsequently confirmed by cleavage of the RNA with the corresponding specific antisense ODNs. In all cases, the cleavage fragments produced by the specific ODNs co-migrated with the corresponding cleavage fragments produced by the library screenings (data not shown). These results demonstrate the utility of triplet-specific semi-randomized ODN libraries in identifying potential target sites.

In vitro cleavage assays were performed with specific ODNs against all sites targeted by ribozymes to investigate whether the sites that were selected by library screening were more accessible than target sites selected by alternative means. The specific ODNs, like those of the semi-randomized libraries, were designed to have the same recognition sequences as a corresponding hypothetical 6+6 armed hammerhead ribozyme. The *in vitro* cleavage assays demonstrated that all the ODNs targeting sites that were selected by library screening resulted in stronger cleavage than the best among the ODNs targeting sites selected by theoretical/structural considerations (Fig. 3B). Furthermore, the strongest cleavage was achieved with the ODN (O-465) corresponding to the GUA site that was cleaved most strongly in the library screenings (Fig. 3A). Thus accessibility data obtained from library screening assays were representative for the hybridization efficiencies of the corresponding unique ODNs. Consequently, major cleavage events do not appear to be substantially influenced by the presence of non-specific ODNs in the library screening. This suggests that neither positive nor negative cooperativity of binding (49) occurs to a significant degree under our assay conditions.

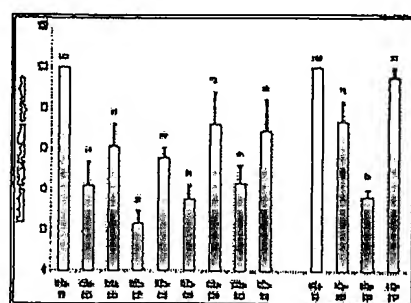
Cellular uptake of ribozyme

FITC-labeled control ribozymes were used to investigate the cellular uptake of ribozyme under the optimized co-transfection conditions by spectroscopy. In order to distinguish between intracellular uptake and membrane association by spectroscopic analysis, uptake experiments were performed in parallel by incubating cells on ice as well as at 37°C. Similar amounts of MRz-FITC and ARz-FITC ribozymes were found to be associated with cells following incubation both on ice (26–28% of total added ribozyme) and at 37°C (58–59% of added ribozyme) (data not shown). The increase in cell-associated fluorescence at the higher temperature, suggests that in excess of 30% of ribozyme has been internalized. The cellular association of ribozyme following incubation on ice was shown to have a linear dependency on ribozyme concentration in the range 0–100 nM (data not shown). This is consistent

with a state of equilibrium between membrane-bound and free (in medium) ribozyme. Assuming that this equilibrium is maintained upon internalization of ribozyme at the permissive temperature, the amount of internalized ribozyme is determined by the formula $I = T(\beta - \alpha)/(T - \alpha)$, where T is the total amount of ribozyme added to cells, and α and β are the measured cell-associations upon incubation on ice and at 37°C, respectively. Net uptake of ribozyme under the given transfection conditions was estimated at 42–45% (data not shown). Co-transfection optimization experiments performed at various w/w ratios (1:1 to 6:1) of lipofectamine to total nucleic acids demonstrated that uptake was only slightly influenced by lipofectamine concentration above a certain threshold (2:1 w/w ratio) that constitutes a molar excess of positive charges (from cationic liposomes) in the complexation mixture. Below this threshold (at a 1:1 w/w ratio, yielding complexes of net negative charge), uptake was severely impaired (data not shown).

Efficacy of ribozymes in co-transfection experiments in HeLa cells

The *in vivo* efficiency of the selected ribozymes was evaluated in a cell culture assay in which 100 nM of ribozyme was co-transfected with a plasmid coding for a PSKH1-luciferase fusion protein. In order to correlate for transfection variability and improve experimental reproducibility, a plasmid coding for *Renilla* luciferase (Rluc) under the control of an EF1 α -promoter was added to the transfection mixture as an internal control. This control proved to be essential as the general expression levels of *Renilla* luciferase varied up to 2–3-fold within the same experiment for different ribozymes (data not shown). Experiments were performed as far as possible with the full complement of ribozymes of identical chemistry so that in each experiment, the efficacies of ribozymes were compared under identical conditions. In each experiment, the relative firefly luciferase activity for all ribozymes was normalized to the levels for the control ribozyme (the expression of which was set at 100%). Normalization was always performed relative to the relevant control ribozyme. The data from experiments with both methylated and allylated ribozymes are summarized in Figure 4. The most effective methylated ribozyme in co-transfection assays proved to be MRz-287, targeting *guc287*. This ribozyme reduced the level (normalized) of luciferase expressed from the reporter gene to ~23% of the control levels, i.e. the levels in cells treated with irrelevant control ribozyme. The second most efficient ribozyme was MRz-519, which resulted in ~65% inhibition of expression. The ribozymes targeting sites *guc118* and *guc548*, exhibited similar inhibitory effects, with residual expression levels just over 40%. At the other end of the spectrum we find the ribozymes targeting *guc712* and *guc539*, which result in only ~30% inhibition.



View larger version (19K):
[in this window]

Figure 4. Lipofectamine-mediated co-transfection of HeLa cells with ribozyme, pTRE-PSK-Luc, and pRLuc as detailed in Materials and Methods. Eight methylated (MRz-) and three allylated (ARz-) PSKH1-specific ribozymes were analyzed together with their respective controls (MRz-TF and ARz-TF). PSKH1-dependent firefly luciferase expression was normalized to *Renilla* luciferase expression for each sample. Normalized expression in cells transfected with control ribozymes within each series was set at 100%. Data are averages of 4–7 independent experiments. Expression levels are indicated above the error bars (+SD) of each column.

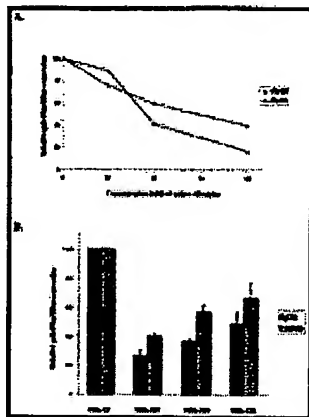
[\[in a new window\]](#)

Three of the ribozymes were synthesized also in the allylated version to evaluate any systematic differences in inhibitory capacity of methylated and allylated ribozymes. The effect of the allylated ribozymes varied over a relatively wide range, resulting in residual reporter gene expression ranging from ~35% for ARz-519 to 95% for ARz-539 (Fig. 4). Comparing the data for these ribozymes with those of their methylated counterparts, we observed the same activity ranking. For both types of modification a ribozyme targeting the *guc519* cleavage site was superior to the ribozymes targeting the two other sites, *gua465* and *guc539*, of which the latter site appeared to be least amenable to cleavage. This constitutes evidence that the observed inhibitory effects of these ribozymes are sequence specific. The difference in activity of methylated and allylated versions of otherwise identical ribozymes varies with the target of the ribozymes. While the two types of ribozymes targeting *guc519* are equally active, the allylated ribozymes targeting *gua465* and *guc539* are less active than their methylated counterparts. The combined data suggest that methylation generally results in superior ribozymes and, further, that the allyl modification seems to be more detrimental to ribozyme efficacy when targeting poorly accessible sites. Reduced efficacy of allylated ribozymes is not due to differences in intracellular uptake, as the methylated and the allylated FITC-labeled ribozyme were internalized to the same level by lipofectamine-mediated transfection (see uptake experiments).

Ribozyme effects depend on ribozyme concentration and target gene promoter strength

The concentration dependence of the ribozyme effect was investigated in co-transfection experiments in which the concentration of the active anti-PSKH1 ribozyme was varied and the total concentration of ribozyme adjusted to 100 nM with control ribozyme. For dose-dependence experiments, the most active allylated (ARz-519) and methylated (MRz-287) ribozymes were selected. These experiments confirmed the dose-dependence of the ribozyme effect (Fig. 5A). We next wanted to investigate whether increasing the steady-state target gene expression would reduce the specific inhibitory effect of ribozymes. Higher expression of the reporter was achieved by placing it under the transcriptional control of the CMV promoter in the pcDNA3 expression plasmid. This resulted in 20-fold enhancement of reporter expression (data not shown) compared to the weaker tetracycline-responsive promoter. In co-transfection experiments, this increase in reporter expression was accompanied by reduced apparent efficiencies of three of the most active PSKH1 ribozymes (Fig. 5B). The average levels of inhibition of target gene expression with the weak promoter were 73, 63 and 51% for ribozymes MRz-287, MRz-519 and MRz-548, respectively. With the strong promoter, yielding the 20-fold higher expression of reporter (measured in control ribozyme-treated cells), their inhibitory activities were 60, 43 and 33%, respectively. Thus the inhibitory effect of ribozymes, as expected, was dependent on the concentrations of both ribozyme and target, but the ribozymes were able to inhibit almost completely the 20-fold increase in the target mRNA.

Figure 5. Dependence of ribozyme inhibitory activity on ribozyme concentration and reporter gene expression. (A) Transfection of HeLa cells with increasing concentrations of ARz-519 or MRz-287



[View larger version](#)
(19K):
[\[in this window\]](#)
[\[in a new window\]](#)

ribozymes. Total concentration of ribozyme is adjusted to 100 nM with control ribozyme (ARz-TF or MRz-TF) and transfections performed after the standard protocol. Data for representative experiments are shown. (B) Parallel co-transfections of HeLa cells with selected ribozymes and different reporter constructs (pTRE-PSK-Luc and pcDNA3-PSK-Luc). Relative expression levels in control cells were 20-fold higher with the pcDNA3-PSK-Luc construct compared to pTRE-PSK-Luc.

Comparing ribozymes selected by MFold secondary structure prediction and *in vitro* accessibility assays

Ribozymes were ranked based on their inhibitory activity in co-transfection assays, with tied ranks given to ribozymes resulting in residual expression levels differing by <3% (Table 2). This classification can be used to assess the relative merits of the two methods of target site selection. The four ribozymes that were selected on the basis of library screening data (MRz-465, MRz-519, MRz-539, MRz-548) were compared to those selected on the basis of structural considerations (MRz-118, MRz-238, MRz-287, MRz-519, MRz-712) by Wilcoxon's rank sum test on unpaired samples (50). The rank sums for the two groups of ribozymes were 18 (average rank = 4.5) and 20 (average rank = 5.0), respectively, which are higher than the critical rank sum of 11 required for a 5% significance level. From these data, ribozyme targets selected on the basis of *in vitro* accessibility are no more susceptible to ribozyme cleavage *in vivo* than those selected by structural prediction or other considerations.

View this table: [Table 2. Rank correlations of ribozyme efficacy with various features of the predicted target site secondary structure](#)
[\[in this window\]](#)
[\[in a new window\]](#)

Correlation of ribozyme efficacy with the predicted target secondary structure

The two methods of target selection both resulted in ribozymes with very diverse inhibitory activities. We therefore decided to investigate whether there was any correlation between predicted structure of target sites and the *in vivo* efficacy of corresponding ribozymes. We approached this by attempting to break down the secondary structures of the target sequences in readily quantifiable parameters. In doing so we have disregarded certain factors that may influence hybridization efficiency but which cannot be easily quantified. Such factors include the positioning of single-stranded stretches within the target sequence and the ability of a specific hammerhead sequence to assume its proper catalytic three-dimensional structure. The presence of so-called free ends, the positioning of single-stranded stretches at

the ends of target sequences, may be expected to be correlated with enhanced antisense binding (12), while elements of strong secondary structure in the neighborhood of the target sequence proper might impede ribozyme folding. We do not imply that these factors are unimportant, merely that as they cannot be easily quantified, other more readily quantifiable parameters should be investigated for predictive value. Three potentially important parameters for target site accessibility were local free energy of folding of the target sequence (11), the size of single-stranded stretches (loops) that may function as 'hooks' for nucleation of duplex formation (51,52), and the length and stability of stems and helices. Stems and helices may need to be opened up for full hybridization of ribozyme to occur and their length and stability may therefore influence hybridization efficiency. Consequently, the target sequences of all ribozymes were decomposed into length of the major loop, stem and helix, while free energies have been calculated both for the target sequence as a whole and for its major helical region (Table 1). Target sequence free energies were calculated for the energetically optimal folding of the RNA, by adding up the energy contributions from all base-pairings, stacking interactions, bulges and various loops that are contained within the target sequences, as indicated by the MFold program. MFold was also used to determine the free energy of the most stable ('limiting') helical region and to estimate the strength of the ribozyme-target hybrid (by folding the corresponding *cis*-acting ribozyme in which catalytic and substrate strands were connected through a 5 nt loop at stem I).

Spearman's rank correlation test (53) was used to investigate the level of correlation of ribozyme efficacy ranking with various features of the predicted secondary structure of the target sequences (Table 2). No correlation was observed with the length of the major loop (Pearson's correlation coefficient $r = 0.16$) or ribozyme-substrate duplex free energy ($r = 0.18$). The correlation with target sequence free energy was weak ($r = 0.35$) and not significant. Ribozyme efficacy was, however, significantly correlated ($P < 0.025$) with both the length of the major base-paired stretch ($r = 0.75$) and the energy of the most stable helix ($r = 0.77$) within the target sequence (Table 2). Correlation was improved ($r = 0.84$) when considering stem length and helix stability together (for this analysis, the rank was taken as the average of the two individual ranks for stem and helix). Correlation coefficients were not significantly affected by ranking ribozyme activity without the use of ties (Table 2).

▷ DISCUSSION

In this study we have analyzed DNA-armed chemically modified hammerhead ribozymes targeting eight GUC and GUA sites selected by two different methods, *in vitro* accessibility assays and MFold prediction. In a co-transfection controlled assay, the ribozymes resulted in residual luciferase reporter gene expression ranging from

23 (MRz-287) to 72% (MRz-539) (Fig. 4). This activity may in part be due to the presence of DNA arms which allow RNase H-mediated cleavage of DNA-RNA hybrids. Target gene expression normalized to the expression of a co-transfected non-target gene allowed the control of any non-target-specific sequence effects of the ribozymes. Non-specific effects are occasionally encountered in association with extended stretches of phosphorothioate (P=S) linkages (19). In an attempt to minimize such effects, no

△	TOP
△	ABSTRACT
△	INTRODUCTION
△	MATERIALS AND METHODS
△	RESULTS
▫	DISCUSSION
▽	REFERENCES

more than two consecutive P=S linkages were incorporated. Finally, ribozymes of different chemistry (with methylated RNA instead of DNA in the arms) against two of the targets (guc519 and guc548) were less efficient in inhibiting target gene expression than their DNA-armed counterparts (unpublished data). In combination, the above data make it unlikely that the observed effects are due to aspects of ribozyme sequence or chemistry unrelated to hybridization-specific activity.

The inhibitory effect of ribozymes was dependent on both the ribozyme concentration (Fig. 5A) and the steady-state expression levels of the reporter gene (Fig. 5B). A reduced inhibitory effect was observed for three selected methylated ribozymes when expressing the reporter gene from a stronger promoter resulting in 20-fold higher expression levels (Fig. 5B). However, the molar amount of target suppression when the stronger promoter was used increased to a degree almost matching the increase in the level of the exogenous target. This interesting fact suggests an enzymatic rather than a stoichiometric effect of the ribozymes on target gene expression. The enzymatic function may derive entirely from the ribozyme or include RNase H activity. It should, however, be noted that the increase in the total level of PSKH1 mRNA by expressing the exogenous transcript from the stronger promoter depends also on the unknown level of the endogenous transcript, which is another target for the ribozyme. The level of suppression achieved with the best ribozyme in this study (77%) is similar to that which has previously been reported for the best of several unmodified ribozymes in a similar luciferase reporter gene co-transfection assay (16). Other comparable studies have reported inhibition levels in the range of 40–80% for a set of 15 ribozymes (54), 50% inhibition with a pair of variously modified ribozymes (29), and 40–55% inhibition for a pair of 2'-F-pyrimidine modified ribozymes with phosphorothioates at both ends (27). Thus our results, achieved with generally more extensively modified ribozymes, compare favorably with previous reports.

We have also attempted to evaluate the relative effects of two commonly employed types of 2'-O-modifications on the activity of ribozymes. The same rank order of ribozymes targeting three selected sites was observed for both methylated and allylated DNA-armed ribozymes (Fig. 4). Allylated ribozymes were generally less efficient in inhibiting reporter gene expression in co-transfection assays compared to their methylated counterparts. The difference in activity of methylated and allylated versions of otherwise identical ribozymes appeared to correlate with the susceptibility of the target to inhibition by ribozyme. While the ribozyme targeting the least accessible of the dually target sites (guc539) apparently was most sensitive to the type of alkylation, ribozyme species of either modification were equally efficient when targeting the most susceptible site (guc519).

Predicted secondary structures of targets selected on the basis of the MFold program were very diverse and included a short stem-loop (guc287), a stem with a large internal loop (guc238), a bulged stem (guc519), and a hairpin structure (guc712), as well as a site near the translation initiation site that was presumed to be relatively unstructured (Fig. 2). *In vitro* accessibility assays with cleavage triplet-specific ODN-libraries identified three accessible GUC sites and one GUA site (Fig. 3A). *In vitro* cleavage assays with specific ODNs against all sites targeted by ribozymes confirmed that the sites that were selected by library screening were indeed more accessible *in vitro* than sites selected by alternative means (Fig. 3B). This confirms the utility of such semi-randomized libraries for identifying the most

accessible sites *in vitro*. However, ribozymes targeting sites selected on the basis of *in vitro* accessibility assays were no more efficient in inhibiting target gene expression in a co-transfection assay than ribozymes targeting sites selected by theoretical means. In fact, the ribozyme targeting the most accessible site *in vitro* (MRz-465) ranked only as the fifth most active (Table 2), while the target site of the best ribozyme (MRz-287), was relatively inaccessible *in vitro* (Fig. 3B). The lack of correlation between *in vitro* accessibility and *in vivo* efficacy data is consistent with previous observations of other researchers (55–58). A possible explanation for the poor correlation is that the target mRNA is folded differently *in vitro* than *in vivo*. The secondary structure of the folded RNA may not be the energetically most stable (fast local folding events may prevent more energetically favorable interactions between distal regions). Structural features that promote hybridization *in vivo* and *in vitro* may also differ. Furthermore, the hybridization efficiency of 13mer antisense ODNs may not be entirely representative for the hybridization characteristics of longer (32mer) hammerhead ribozymes with varying degrees of secondary structure of their own. Finally, the generation of higher-order mRNA structures and modulation or masking of the mRNA secondary structure by RNA-binding proteins (9,10) or protein complexes (ribosomes) (59) *in vivo* may influence accessibility of target sites, although these factors would also tend to rule out the applicability of secondary structure predictions for target site selection. Recent data suggest that performing *in vitro* accessibility assays on an endogenous transcript in a protein environment (cell extracts) may improve the accuracy of the predictions (14,22).

Although *in vitro* accessibility assays proved to be of limited predictive value for the *in vivo* situation, correlative studies suggested that secondary structure predictions might have some merit. Significant correlation was found between ribozyme efficacy and the presence of short stems and energetically unstable helices within the ribozyme target sequence (Table 2). Ranking of ribozymes according to these two criteria, the relative efficacies of ribozymes were predicted nearly perfectly, the only significant discrepancy being the ribozyme targeting *gua465*. As well as being most accessible *in vitro*, this site also has a secondary structure that according to the above criteria should make it a significantly better target site than observed here. Other factors, such as the lack of single-stranded bases near the ends of the target sequence or a prohibitive environment for ribozyme folding, may explain the results. Notwithstanding this discrepancy, there is an apparently clear correlation between ribozyme efficacy and predicted target sequence secondary structure. Furthermore, the combined data from this study suggest that the previously reported correlation of ribozyme efficacy with local folding potential (11) may be incidental. Our data suggest that high target sequence free energy alone may not be sufficient for efficient ribozyme targeting. We propose that the above correlation is a consequence of the need to have some unpaired regions to facilitate fast nucleation of duplex formation (51,52), combined with short base-paired regions and helices of low stability that easily open up. Fulfillment of these criteria will in many cases result in a low local folding potential (high free energy) for the target site.

A recent study on the effect of varying RNA secondary structure on the efficiency of specific antisense ODNs concluded that target sequences located within regions designed to be unstructured were most effective, while targets within stable stem-loop structures were ineffective (13). Recent reports by Patzel and co-workers (12,14) described a theoretical approach for antisense ODN target site selection based on the prediction of large single-stranded stretches (loops) by MFold. Our data do not support a correlation

of ribozyme activity with the length of loops, possibly because all our targets had shorter predicted loops than recommended in the above studies. Our hypothesis and the conclusions of the above studies are, however, not mutually exclusive. Target sites containing very large loops will have a good probability of also containing short helical regions, which we propose to be the limiting factor. In fact, applying our hypothesis of target site evaluation to three previously well characterized target sites (t351, t398, t498) (14,22) within mRNA for murine DNA methyl transferase, the same rank susceptibility of targets, in perfect accordance with actual inhibition data, is predicted by both theories.

In addition to the structure of the target site, the composition of the target sequence may also be of some importance. Sequences with a high G+C content will hybridize more efficiently with the complementary arms of their ribozymes and possibly increase the efficacy of the ribozyme. Although a general correlation between hybrid stability and ribozyme efficacy was not supported by our data (Table 2), it is worth noting that the most efficient ribozyme, MRz-287, has a substantially higher affinity for its target sequence than the other ribozymes (Table 1) due to an unusually high G+C content (13 out of 16 nt). In conclusion, the target sequence of MRz-287 represents the proposed desirable structural features for a good ribozyme/antisense target site. The target sequence consists of alternating short stretches of paired and unpaired bases, which limit stems and helical regions to no more than 3 bp. All other target sequences fold into secondary structures containing longer and more stable helical regions. One study has reported that hybridization accessibility for hammerhead ribozymes is correlated with the presence of unpaired bases near the cleavage triplet (60). Although our data do not suggest this to be a critical requirement for *in vivo* activity, the above criterium is also fulfilled for the *guc287* target site, as the longest single-stranded stretch is situated around the cleavage triplet and includes the base preceding the scissile bond (Fig. 2). All of the above mentioned features of *guc287* add up to a very effective ribozyme target site, in good agreement with its observed inhibitory capacity in HeLa cells.

In conclusion, our study indicates that there is a poor correlation between the apparent *in vivo* accessibility of a target and its accessibility in a completely cell-free *in vitro* assay as performed here. Thus such assays appear to be of limited value even for a preliminary selection of target sites. However, predictions by the MFold program suggest a correlation of certain features of the predicted secondary structures of target sequences, helical stability in particular, with ribozyme efficacy. The generality of these findings will, however, need to be investigated in an alternate test system. If these correlations should be confirmed, this would represent a significant improvement in the preliminary selection of candidate ribozymes. Ultimately, however, an empirical cell-based assay will still need to be performed to select the best of these candidates.

▷ ACKNOWLEDGEMENTS

M.A. is a research fellow of the Norwegian Cancer Society. This work was supported by grants to H.P. from the Norwegian Cancer Society and the Norwegian Research Council.

▷

FOOTNOTES

* To whom correspondence should be addressed. Tel: +47 2295 8755; Fax: +47 2269 4130; Email: hans.prydz@biotek.uio.no

► REFERENCES

▲ TOP
▲ ABSTRACT
▲ INTRODUCTION
▲ MATERIALS AND METHODS
▲ RESULTS
▲ DISCUSSION
▪ REFERENCES

1 Amarzguioui, M. and Prydz, H. (1998) *Cell. Mol. Life Sci.*, **54**, 1175–1202. [[ISI](#)][[Medline](#)]

2 Ludwig, J., Blaschke, M. and Sproat, B.S. (1998) *Nucleic Acids Res.*, **26**, 2279–2285. [[Abstract/Free Full Text](#)]

3 Vaish, N.K., Heaton, P.A., Fedorova, O. and Eckstein, F. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2158–2161. [[Abstract/Free Full Text](#)]

4 Milner, N., Mir, K.U. and Southern, E.M. (1997) *Nat. Biotechnol.*, **15**, 537–541. [[ISI](#)][[Medline](#)]

5 Ho, S.P., Bao, Y., Leshner, T., Malhotra, R., Ma, L.Y., Fluharty, S.J. and Sakai, R.R. (1998) *Nat. Biotechnol.*, **16**, 59–63. [[ISI](#)][[Medline](#)]

6 Brede, G. (2000) EMBL accession no. AJ272212.

7 Zuker, M., Mathews, D.H. and Turner, D.H. (1999) In Barciszewski, J. and Clark, B.F.C. (eds), *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–43.

8 Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) *J. Mol. Biol.*, **288**, 911–940. [[ISI](#)][[Medline](#)]

9 Tsuchihashi, Z., Khosla, M. and Herschlag, D. (1993) *Science*, **267**, 99–102.

10 Sioud, M. (1994) *J. Mol. Biol.*, **242**, 619–629. [[ISI](#)][[Medline](#)]

11 Sczakiel, G., Homann, M. and Rittner, K. (1993) *Antisense Res. Dev.*, **3**, 45–52. [[Medline](#)]

12 Patzel, V., Steidl, U., Kronenwett, R., Haas, R. and Sczakiel, G. (1999) *Nucleic Acids Res.*, **27**, 4328–4334. [[Abstract/Free Full Text](#)]

13 Vickers, T.A., Wyatt, J.R. and Freier, S.M. (2000) *Nucleic Acids Res.*, **28**, 1340–1347. [[Abstract/Free Full Text](#)]

14 Scherr, M., Rossi, J.J., Sczakiel, G. and Patzel, V. (2000) *Nucleic Acids Res.*, **28**, 2455–2461. [[Abstract/Free Full Text](#)]

- 15 Shimayama,T., Nishikawa,S. and Taira,K. (1993) *Nucleic Acids Res. Symp. Ser.*, **29**, 177–178.
- 16 Sakamoto,N., Wu,C.H. and Wu,G.Y. (1996) *J. Clin. Invest.*, **98**, 2720–2728.
[Abstract/Free Full Text]
- 17 Szymanski,M., Fürste,J.P., Barciszewska,M.Z., Erdmann,V.A. and Barciszewski,J. (1997) *Biochem. Mol. Biol. Int.*, **41**, 439–447.[ISI][Medline]
- 18 Mir,K.U. and Southern,E.M. (1999) *Nat. Biotechnol.*, **17**, 788–792.[ISI][Medline]
- 19 Jarvis,T.C., Wincott,F.E., Alby,L.J., McSwiggen,J.A., Beigelman,L., Gustofson,J., DiRenzo,A., Levy,K., Arthur,M., Matulic-Adamic,J., Karpeisky,A., Gonzalez,C., Woolf,T.M., Usman,N. and Stinchcomb,D.T. (1996) *J. Biol. Chem.*, **271**, 29107–29112.
[Abstract/Free Full Text]
- 20 Ho,S.P., Britton,D.H., Stone,B.A., Behrens,D.L., Leffet,L.M., Hobbs,F.W., Miller,J.A. and Trainor,G.L. (1996) *Nucleic Acids Res.*, **24**, 1901–1907.[Abstract/Free Full Text]
- 21 Birikh,K.R., Berlin,Y.A., Soreq,H. and Eckstein,F. (1997) *RNA*, **3**, 429–437.[Abstract]
- 22 Scherr,M. and Rossi,J.J. (1998) *Nucleic Acids Res.*, **26**, 5079–5085.[Abstract/Free Full Text]
- 23 Matveeva,O., Felden,B., Audlin,S., Gesteland,R.F. and Atkins,J.F. (1997) *Nucleic Acids Res.*, **25**, 5010–5016.[Abstract/Free Full Text]
- 24 Zoumadakis,M. and Tabler,M. (1995) *Nucleic Acids Res.*, **23**, 1192–1196.[Abstract]
- 25 Lieber,A. and Strauss,M. (1995) *Mol. Cell. Biol.*, **15**, 540–551.[Abstract]
- 26 Hertel,K.J., Pardi,A., Uhlenbeck,O.C., Koizumi,M., Ohtsuka,E., Uesugi,S., Cedergren,R., Eckstein,F., Gerlach,W.L., Hodgson,R. et al. (1992) *Nucleic Acids Res.*, **20**, 3252.[ISI][Medline]
- 27 Scherr,M., Grez,M., Ganser,A. and Engels,J.W. (1997) *J. Biol. Chem.*, **272**, 14304–14313.
[Abstract/Free Full Text]
- 28 Parry,T.J., Cushman,C., Gallegos,A.M., Agrawal,A.B., Richardson,M., Andrews,L.E., Maloney,L., Mokler,V.R., Wincott,F.E. and Pavco,P.A. (1999) *Nucleic Acids Res.*, **27**, 2569–2577.[Abstract/Free Full Text]
- 29 Bramlage,B., Alefelder,S., Marshall,P. and Eckstein,F. (1999) *Nucleic Acids Res.*, **27**, 3159–3167.[Abstract/Free Full Text]
- 30 Tuschl,T. and Eckstein,F. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6991–6994.[Abstract]
- 31 Beigelman,L., Karpeisky,A., Matulic-Adamic,J., Haeberli,P., Sweedler,D. and Usman,N. (1995) *Nucleic Acids Res.*, **23**, 4434–4442.[ISI][Medline]
- 32 Lyngstadaas,S.P., Risnes,S., Sproat,B.S., Thrane,P.S. and Prydz,H.P. (1995) *EMBO J.*, **14**, 5224–5229.[Abstract]

- 33 Gu,J.L., Veerapanane,D., Rossi,J., Natarajan,R., Thomas,L. and Nadler,J. (1995) *Circ. Res.*, **77**, 14–20.[[Abstract/Free Full Text](#)]
- 34 Wang,Q., Mullah,B., Hansen,C., Asundi,J. and Robishaw,J.D. (1997) *J. Biol. Chem.*, **272**, 26040–26048.[[Abstract/Free Full Text](#)]
- 35 Flory,C.M., Pavco,P.A., Jarvis,T.C., Lesch,M.E., Wincott,F.E., Beigelman,L., Hunt,S.W.,III and Schrier,D.J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 754–758.[[Abstract/Free Full Text](#)]
- 36 Sproat,B.S., Rupp,T., Menhardt,N., Keane,D. and Beijer,B. (1999) *Nucleic Acids Res.*, **27**, 1950–1955.[[Free Full Text](#)]
- 37 Hendry,P., McCall,M.J., Santiago,F.S. and Jennings,P.A. (1992) *Nucleic Acids Res.*, **20**, 5737–5741.[[Abstract](#)]
- 38 Shimayama,T., Nishikawa,F., Nishikawa,S. and Taira,K. (1993) *Nucleic Acids Res.*, **21**, 2605–2611.[[Abstract](#)]
- 39 Shimayama,T. (1994) *Gene*, **149**, 41–46.[[ISI](#)][[Medline](#)]
- 40 Heidenreich,O., Xu,X., Swiderski,P., Rossi,J.J. and Nerenberg,M. (1996) *Antisense Nucleic Acid Drug Dev.*, **6**, 111–118.[[ISI](#)][[Medline](#)]
- 41 Sinha,N.D., Biernat,J., McManus,J. and Köster,H. (1984) *Nucleic Acids Res.*, **12**, 4539–4557. [[Abstract](#)]
- 42 Usman,N., Ogilvie,K.K., Jiang,M.-Y. and Cedergren,R.J. (1987) *J. Am. Chem. Soc.*, **109**, 7845–7854.[[ISI](#)]
- 43 Green,R., Szostak,J.W., Benner,S.A., Rich,A. and Usman,N. (1991) *Nucleic Acids Res.*, **19**, 4161–4166.[[Abstract](#)]
- 44 Polushin,N.N., Pashkova,I.N. and Efimov,V.A. (1991) *Nucleic Acids Res. Symp. Ser.*, **24**, 49.
- 45 Polushin,N.N., Morocho,A.M., Chen,B. and Cohen,J.S. (1994) *Nucleic Acids Res.*, **22**, 639–645.[[Abstract](#)]
- 46 Vinayak,R., Andrus,A. and Hampel,A. (1995) *Biomed. Pept. Proteins Nucleic Acids*, **1**, 227–230.[[Medline](#)]
- 47 Wang,F.S., Kobayashi,H., Liang,K.W., Holland,J.F. and Ohnuma,T. (1999) *Hum. Gene Ther.*, **10**, 1185–1195.[[ISI](#)][[Medline](#)]
- 48 Kobayashi,H., Takemura,Y., Wang,F.S., Oka,T. and Ohnuma,T. (1999) *Int. J. Cancer*, **81**, 944–950.[[ISI](#)][[Medline](#)]
- 49 Jankowsky,E. and Schwenzer,B. (1996) *Biochemistry*, **35**, 15313–15321.[[ISI](#)][[Medline](#)]
- 50 Wilcoxon,F. (1945) *Biomet. Bull.*, **6**, 80.

51 Homann,M., Rittner,K. and Sczakiel,G. (1993) *J. Mol. Biol.*, **233**, 7–15.[ISI][Medline]

52 Skripkin,E., Paillart,J.C., Marquet,R., Blumenfeld,M., Ehresmann,B. and Ehresmann,C. (1996) *J. Biol. Chem.*, **271**, 28812–28817.[Abstract/Free Full Text]

53 Altman,D.G. (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London, UK, pp. 293–295.

54 Macejak,D.G., Jensen,K.L., Jamison,S.F., Domenico,K., Roberts,E.C., Chaudhary,N., von Carlowitz,I., Bellon,L., Tong,M.J., Conrad,A., Pavco,P.A. and Blatt,L.M. (2000) *Hepatology*, **31**, 769–776.[ISI][Medline]

55 L'Huillier,P.J., Davis,S.R. and Bellamy,A.R. (1992) *EMBO J.*, **11**, 4411–4418.[Abstract]

56 Beck,J. and Nassal,M. (1995) *Nucleic Acids Res.*, **23**, 4954–4962.[Abstract]


57 Domi,A., Beaud,G. and Favre,A. (1996) *Biochimie*, **78**, 654–662.[ISI][Medline]


58 Ramezani,A. and Joshi,S. (1996) *Antisense Nucleic Acid Drug. Dev.*, **6**, 229–235.[ISI][Medline]

59 Chen,H., Ferbeyre,G. and Cedergren,R. (1997) *Nat. Biotechnol.*, **15**, 432–435.[ISI][Medline]

60 Campbell,T.B., McDonald,C.K. and Hagen,M. (1997) *Nucleic Acids Res.*, **25**, 4985–4993. [Abstract/Free Full Text]

This article has been cited by other articles:

 **The Journal of Gerontology: BIOLOGICAL SCIENCES AND MEDICAL SCIENCES** [HOME](#)
C. C. Deocaris, S. C. Kaul, K. Taira, and R. Wadhwa
Emerging Technologies: Trendy RNA Tools for Aging Research
J. Gerontol. A Biol. Sci. Med. Sci., August 1, 2004; **59**(8): B771 - B783.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

 **Nucleic Acids Research** [HOME](#)
D. Boden, O. Pusch, F. Lee, L. Tucker, P. R. Shank, and B. Ramratnam
Promoter choice affects the potency of HIV-1 specific RNA interference
Nucleic Acids Res., September 1, 2003; **31**(17): 5033 - 5038.
[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)

Nucleic Acids Research [HOME](#)
R. Kretschmer-Kazemi Far and G. Sczakiel
The activity of siRNA in mammalian cells is related to structural target accessibility: a comparison with antisense oligonucleotides



Nucleic Acids Res., August 1, 2003; 31(15): 4417 - 4424.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Nucleic Acids Research

[▶ HOME](#)

T. Holen, M. Amarzguloul, E. Babale, and H. Prydz
Similar behaviour of single-strand and double-strand siRNAs suggests they act through a common RNAi pathway

Nucleic Acids Res., May 1, 2003; 31(9): 2401 - 2407.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Nucleic Acids Research

[▶ HOME](#)

M. Amarzguloui, T. Holen, E. Babale, and H. Prydz
Tolerance for mutations and chemical modifications in a siRNA

Nucleic Acids Res., January 15, 2003; 31(2): 589 - 595.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Nucleic Acids Research

[▶ HOME](#)

L. J. Bergeron and J.-P. Perreault
Development and comparison of procedures for the selection of delta ribozyme cleavage sites within the hepatitis B virus

Nucleic Acids Res., November 1, 2002; 30(21): 4682 - 4691.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Journal of Virology

[▶ HOME](#)

R. M. Smith, C. M. Walton, C. H. Wu, and G. Y. Wu
Secondary Structure and Hybridization Accessibility of Hepatitis C Virus 3'-Terminal Sequences

J. Virol., August 28, 2002; 76(19): 9563 - 9574.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Nucleic Acids Research

[▶ HOME](#)

T. Holen, M. Amarzguloui, M. T. Wilger, E. Babaie, and H. Prydz
Positional effects of short interfering RNAs targeting the human coagulation trigger Tissue Factor

Nucleic Acids Res., April 15, 2002; 30(8): 1757 - 1766.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



JBC Online

[▶ HOME](#)

J. Kurreck, B. Bieber, R. Jahnel, and V. A. Erdmann
Comparative Study of DNA Enzymes and Ribozymes against the Same Full-length Messenger RNA of the Vanilloid Receptor Subtype I

J. Biol. Chem., March 1, 2002; 277(9): 7099 - 7107.

[\[Abstract\]](#) [\[Full Text\]](#) [\[PDF\]](#)



Nucleic Acids Research

[HOME](#)

A. A. Mir, T. J. Lockett, and P. Hendry

Identifying ribozyme-accessible sites using NUH triplet-targeting gapmers

Nucleic Acids Res., May 1, 2001; 29(9): 1906 - 1914.

[\[Abstract\]](#) [\[Full Text\]](#)

This Article

- ▶ [Abstract](#) FREE
- ▶ [Print PDF \(331K\)](#) FREE
- ▶ [Alert me when this article is cited](#)
- ▶ [Alert me if a correction is posted](#)

Services

- ▶ [Email this article to a friend](#)
- ▶ [Similar articles in this journal](#)
- ▶ [Similar articles in ISI Web of Science](#)
- ▶ [Similar articles in PubMed](#)
- ▶ [Download to citation manager](#)
- ▶ [Search for citing articles in: ISI Web of Science \(34\)](#)
- ▶ [Commercial Re-use Guidelines for Open Access NAR Content](#)

PubMed

- ▶ [PubMed Citation](#)
- ▶ [Articles by Amarzguioui, M.](#)
- ▶ [Articles by Prydz, H.](#)

ITEM 7

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: KHVOROVA Examiner: to be assigned
Serial No.: 10/714333 Group Art Unit: 1646
Filed: November 14, 2003
For: Functional and Hyperfunctional siRNA
Customer No.: 23719

Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

March 7, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

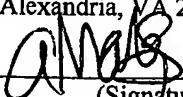
INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents:

In accordance with 37 C.F.R. §§ 1.56 and 1.97 through 1.98, Applicants wish to make known to the Patent and Trademark Office the references set forth on the attached form PTO-1449 (copies of cited references are enclosed). As to any reference supplied, Applicants do not admit that it is "prior art" under 35 U.S.C. §§ 102 or 103, and specifically reserves the right to traverse or antedate any such reference, as by a showing under 37 C.F.R. § 1.131 or other method. Although the aforesaid references are made known to the Patent and Trademark Office in compliance with Applicants' duty to disclose all information of which they are aware and believe relevant to the examination of the above-identified application, Applicants believe that their invention is patentable.

Certificate of Mailing Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited on the date shown below with the United States Postal Service as first class mail with sufficient postage in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.


(Signature)

Angel Matos
(Printed Name of Person Signing Certificate)

3/7/05
(Date)

I. U.S. PATENT DOCUMENTS

<u>Patent No.</u>	<u>Date Issued</u>	<u>Title</u>
US 2002/0150945 A1	October 17, 2002	Method for Making Polynucleotide Libraries, Polynucleotide Arrays, and Cell Libraries for High-Throughput Genomics

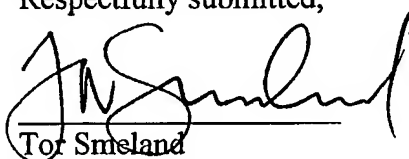
III. NON PATENT PUBLICATIONS

AMARZGUIOUI *et al.*, Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes. 2000 Oxford University Press, Nucleic Acids Research, 2000, Vol. 28, No. 21 4113-4124.

KASIF *et al.* A computational framework for optimal masking in the synthesis of oligonucleotide microarrays. 2002 Oxford University Press, Nucleic Acids Research, 2002, Vol. 30 No.20 *e106*.

Because no action has been taken on the merits, Applicants submit that no fee is due at this time. However, if a fee is deemed necessary, please charge Deposit Account No. 11-0171.

Respectfully submitted,



Tor Smeland

Registration No.: 43,131
Attorney for Applicant

Kalow & Springut LLP
(212) 813-1600

Substitute for form 1449/PTO

PTO/SB/08A (08-03)

Approved for use through 07/31/2006. OMB 0651-0031

U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

U.S. Patent and Trademark Office, U.S. DEPARTMENT OF COMMERCE
 to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT

(Use as many sheets as necessary)

Sheet

1

of

2

Complete if Known

Application Number

10/714333

Filing Date

November 14, 2003

First Named Inventor

KHVOROVA *et al.*

Art Unit

1646

Examiner Name

to be assigned

Attorney Docket Number

13499 US

U. S. PATENT DOCUMENTS

[illegible]

FOREIGN PATENT DOCUMENTS

[illegible]

**Examiner
Signature**

Date _____

Considered

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant. ¹ Applicant's unique citation designation number (optional). ² See Kinds Codes of USPTO Patent Documents at www.uspto.gov or MPEP 901.04. ³ Enter Office that issued the document, by the two-letter code (WIPO Standard ST.3). ⁴ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁵ Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. ⁶ Applicant is to place a check mark here if English language Translation is attached.

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. **DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Substitute for form 1449/PTO

**INFORMATION DISCLOSURE
STATEMENT BY APPLICANT**

(Use as many sheets as necessary)

Complete if Known

Application Number	10/714333
Filing Date	November 14, 2003
First Named Inventor	KHVOROVA <i>et al.</i>
Art Unit	1646
Examiner Name	to be assigned
Attorney Docket Number	13499 US

Sheet	2	of	2
-------	---	----	---

NON PATENT LITERATURE DOCUMENTS

Examiner Initials*	Cite No. ¹	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	T ²
		Amarzguioul, Mohammed <i>et al.</i> Secondary structure prediction and <i>in vitro</i> accessibility of mRNA as tools in the selection of target sites for ribozymes 2000 Oxford University Press, Nucleic Acids Research, 2000, Vol. 28, No.21 pages 4113-4124	
		Kasif, Simon <i>et al.</i> A computational framework for optimal masking in the synthesis of oligonucleotide microarrays 2002 Oxford University Press, Nucleic Acid Research, 2002, Vol.30 No.20 e106	
		International Search Report for International Application No. PCT/US03/36787	

Examiner Signature		Date Considered	
--------------------	--	-----------------	--

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ Applicant's unique citation designation number (optional). ² Applicant is to place a check mark here if English language Translation is attached. This collection of information is required by 37 CFR 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 2 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 (1-800-786-9199) and select option 2.

Secondary structure prediction and *in vitro* accessibility of mRNA as tools in the selection of target sites for ribozymes

Mohammed Amarzguioui, Gaute Brede, Eshrat Babale, Morten Grøtli¹, Brian Sproat² and Hans Prydz*

The Biotechnology Centre of Oslo, University of Oslo, Gaustadalleen 21, 0349 Oslo, Norway, ¹Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark and ²Institut für Organische Chemie, Universität Göttingen, Tammann strasse 2, 37077 Göttingen, Germany

Received July 27, 2000; Revised and Accepted September 18, 2000

DDBJ/EMBL/GenBank accession no. AJ272212

ABSTRACT

We have investigated the relative merits of two commonly used methods for target site selection for ribozymes: secondary structure prediction (Mfold program) and *in vitro* accessibility assays. A total of eight methylated ribozymes with DNA arms were synthesized and analyzed in a transient co-transfection assay in HeLa cells. Residual expression levels ranging from 23 to 72% were obtained with anti-PSKH1 ribozymes compared to cells transfected with an irrelevant control ribozyme. Ribozyme efficacy depended on both ribozyme concentration and the steady state expression levels of the target mRNA. Alkylated ribozymes against a subset of the target sites generally displayed poorer efficacy than their methylated counterparts. This effect appeared to be influenced by *in vivo* accessibility of the target site. Ribozymes designed on the basis of either selection method displayed a wide range of efficacies with no significant differences in the average activities of the two groups of ribozymes. While *in vitro* accessibility assays had limited predictive power, there was a significant correlation between certain features of the predicted secondary structure of the target sequence and the efficacy of the corresponding ribozyme. Specifically, ribozyme efficacy appeared to be positively correlated with the presence of short stem regions and helices of low stability within their target sequences. There were no correlations with predicted free energy or loop length.

INTRODUCTION

Hammerhead ribozymes are potentially powerful tools for sequence-specific inhibition of target gene expression (1). Their intrinsic cleavage activity makes them theoretically superior to traditional antisense oligodeoxynucleotides

(ODNs) in terms of inhibitory capacity. Recent advances (1–3) have extended the range of targets so far that virtually any limited stretch of RNA is now likely to contain a useful target. However, other problems, including methods of delivery and target site selection, remain to be solved. The latter, in particular, appears to be a critical step in the design of antisense or ribozyme molecules for suppression of target gene expression, as there may exist only a few sites within any mRNA that are accessible to hybridization (4,5). We have compared two commonly employed methods for the rational selection of target sites, secondary structure prediction and *in vitro* accessibility assays. As target we have used the mRNA for a novel human protein kinase, PSKH1 (6).

RNA secondary structure prediction algorithms based mainly on energy minimization have intrinsic limitations, although subject to continuous improvement (7,8). An established theory for selection of target sites based on predicted structure is also required. The possible modulation of RNA structures by protein binding *in vivo* (9,10) cannot yet be modeled and has so far limited their use in designing antisense ODNs or ribozymes. A positive correlation between the inhibitory effect of an antisense RNA and low local folding potential has been noted (high ΔG) (11). Together with more recent data (12–14), this suggests that extended single-stranded or unstructured regions may be the best targets for antisense ODNs. However, although ribozymes targeting predicted loop regions have proved efficacious in some cases (15,16), there are also examples of failures (17). A recent systematic analysis of the hybridization of tRNA-Phe to a set of complementary ODNs determined that all high-yield heteroduplexes involved RNA sequences forming both double-stranded stems and single-stranded regions, and that bases of the latter regions were often stacked onto the stems (18). This suggests the requirement for at least some degree of helical conformation in the secondary structures of favorable targets.

The accessibility of different stretches of the mRNA for hybridization with short antisense ODNs may be determined *in vitro* by RNase H-mediated cleavage of the RNA at regions where the ODNs hybridize to the target transcript (14,19–22). Screening of a large set of ODNs targeting potential ribozyme

*To whom correspondence should be addressed. Tel: +47 2295 8755; Fax: +47 2269 4130; Email: hans.prydz@biotek.uio.no

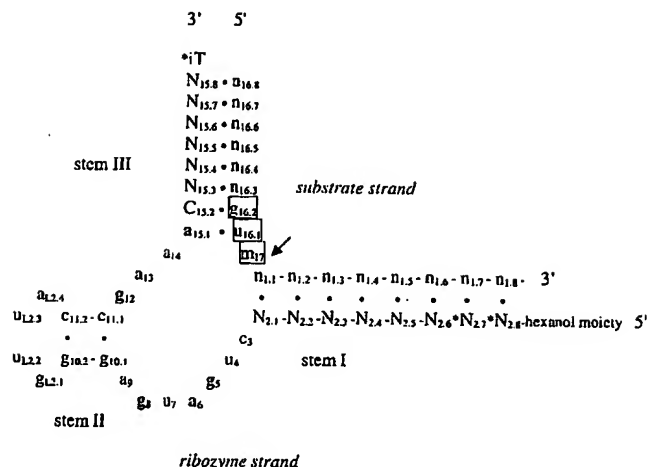


Figure 1. Schematic of the hammerhead ribozyme hybridizing to its target mRNA. The arrow indicates the position of cleavage in the mRNA. Numbering is according to the nomenclature of Hertel *et al.* (26). Unmodified ribonucleotides are in bold lower case, 2'-O-alkylated ribonucleotides in plain lower case, and deoxynucleotides in upper case. Phosphorothioates are indicated by asterisk, while iT denotes an inverted 3'-3' thymidine. Bases of the GUM target triplet, where M is C or A, are boxed.

cleavage sites has been employed to select the most promising sites for ribozyme targeting (19). Accessibility assays with specific ODNs have also been performed in cell extracts on endogenous transcripts in order to better approximate the *in vivo* situation (14,22). The most accessible sites within any target RNA may be selected by performing the RNase H-assay with a randomized set of ODNs (20,21) or with a target-specific set of cDNA fragments prepared by partial DNase I-digestion (23). ODN libraries in previous studies (20,21) have been more or less completely randomized. We wanted to identify specific sites amenable to cleavage with standard hammerhead ribozymes, and to restrict the analysis to those triplets that are cleaved most efficiently. Since the hammerhead ribozyme seems to prefer a purine in the first position of the cleavage triplet and a C or an A in the third position (24), we decided to concentrate on the triplets GUC, GUA, AUC and AUA in our analysis. Four ODN libraries, each specific for one of these cleavage triplets, were synthesized and used to screen *in vitro* transcribed PSKH1 RNA for accessible sites.

Based on a study of the most active ribozymes expressed *in vivo* from a randomized library (25), the ribozymes used in the present study were designed to have symmetric 8+8 nt arms (Fig. 1) (26). Chemically synthesized ribozymes of similar arm lengths have subsequently been successfully employed (19,27–29). The stem II structure (Fig. 1) was chosen to be 2bp long (30). The activity of such truncated ribozymes in cell culture has been demonstrated (27,29). For increased nuclease stability of the ribozymes, we have retained unmodified ribonucleotides in only five catalytically important positions (31,32). Deoxyribonucleotides were used in the flanking arms (33,34) and 2'-O-alkylated ribonucleotides in the core and stem-loop II (31,32,35) (Fig. 1). Nuclease stability of ribozymes was increased further by an inverted thymidine at the 3'-end (28,29,32) and by a hexanol moiety at the extreme 5'-end (36). Ribozymes used *in vivo* frequently include short stretches of

phosphorothioate linkages at the 5'-end (28,29), 3'-end (34) or both (27) for stabilization against exonucleases. We included two phosphorothioates at the 5'-end and one at the 3'-end. DNA nucleotides in the arms have been reported to increase catalytic efficiencies of ribozymes *in vitro*, most likely due to increased product dissociation rates (37–39). DNA-armed ribozymes may recruit RNase H activity upon hybridization with the target RNA and thus enhance their apparent activity (40). We are here primarily interested in the accessibility of cleavage sites. Anything that reduces the importance of intrinsic ribozyme cleavage activity and increases the importance of target-specific inhibition of expression is desirable.

Cellular delivery of ribozymes is commonly accomplished with various cationic liposome formulations (16,19,27–29). We have used the cationic liposome reagent lipofectamine to co-transfect HeLa cells with a mixture of ribozyme, a firefly luciferase reporter gene construct containing the complete coding cDNA of the target PSKH1, and a *Renilla* luciferase-encoding plasmid serving as an internal transfection control. We have succeeded in constructing a ribozyme against PSKH1 mRNA which reduced the activity of the corresponding reporter gene to 20–25% in a concentration dependent manner. A correlation is observed with certain features of the predicted secondary structure of the target mRNA.

MATERIALS AND METHODS

All restriction enzymes were from New England Biolabs.

Synthesis and purification of hammerhead ribozymes

Automated RNA and DNA synthesis was carried out on an Applied Biosystems model 394 DNA/RNA synthesizer. 2'-O-Alkyl ribozymes containing five unmodified purine ribonucleotides were synthesized by solid phase β -cyanoethyl phosphoramidite chemistry (41), using the 2'-O-*tert*-butyldimethylsilyl protection strategy for the ribonucleotides (42,43). Syntheses were performed on controlled pore glass bearing an inverted thymidine linkage (Glen Research). A lipophilic capture tag, 1-[Diisopropyl(DL- α -tocopheryloxy)silyl]oxy-6-(2-cyanoethyl *N,N*-diisopropylamino-phosphinoxy)hexane, was added at the 5'-end of the oligomer as described (36). Cleavage from the support and release of all base labile protecting groups (44,45), reverse phase HPLC purification (Pharmacia Source 5RPC 10/10 column, using a flow-rate of 1 ml/min, or a μ Bondapac C-18 column), lyophilization, desilylation (46), butanol-precipitation and counter-ion exchange with NaClO₄ were performed essentially as described (36). Ribozymes, retaining a 5' hexanol-linker, were desalted (NAP-10 columns, Pharmacia) and quantified by UV spectroscopy. Molar extinction coefficients were calculated based on the nearest-neighbor method (Biopolymer Calculator at <http://paris.chem.yale.edu>). Exact molecular weights were calculated. Ribozymes were controlled by denaturing 15% polyacrylamide gel electrophoresis prior to their application in cell culture experiments.

Sequence and modification of hammerhead ribozymes

PSKH1-specific 2'-O-methylated ribozymes (MRz) were designed targeting a total of eight sites. Allylated ribozymes (ARz) targeting a subset of these sites were also synthesized in order to evaluate the relative effects of the two types of modification on the efficacy of ribozymes. For both types of

ribozymes, controls having the same chemical composition and length of hybridizing arms were designed targeting an irrelevant mRNA (human tissue factor). Ribozymes are named after their type of modification and cleavage position. Thus MRz-519 is a methylated ribozyme cleaving after *guc519*, while ARz-519 is the corresponding allylated ribozyme. The respective control ribozymes are MRz-TF and ARz-TF. The unique sequences (flanking arms) of the ribozymes were as follows (the conserved sequence of the core is indicated by <core> for all but the first ribozyme sequence): MRz-TF/ARz-TF, A*A*T-C-T-C-C-T-c-u-g-a-u-g-a-g-g-u-u-a-c-g-a-a-C-T-T-A-G-T*G-iT; MRz-118, T*C*G-G-G-A-A-G- <core> -a-C-C-T-T-G-C*T-iT; MRz-238, C*C*G-G-C-T-T-T <core> -a-C-A-G-G-G-C*C-iT; MRz-287, A*G*T-C-G-G-G-G- <core> -a-C-C-G-G-G-G*C-iT; MRz-465/ARz-465, T*G*A-T-G-G-C-A- <core> -a-C-G-G-C-T-G*C-iT; MRz-519/ARz-519, G*C*A-G-C-T-C-C- <core> -a-C-T-C-A-C-A*C-iT; MRz-539/ARz-539, A*C*G-C-A-C-C-C- <core> -a-C-G-C-A-G-C*A-iT; MRz-548, G*T*T-G-G-C-A-T- <core> -a-C-G-C-A-C-C*C-iT; MRz-712, A*G*A-T-A-C-C-G- <core> -a-C-G-C-C-A-T*C-iT. Modified and unmodified ribonucleotides are in plain and bold lower case, respectively, while deoxyribonucleotides are in upper case. An asterisk denotes a phosphorothioate linkage, while iT indicates an inverted 3'-3' thymidine (Fig. 1).

Specific and semi-randomized DNA ODNs

Semi-randomized 13mer antisense ODNs specific for each of the target triplets GUC, GUA, AUC and AUA were synthesized and HPLC-purified. The ODNs were designed to mimic the hybridizing arms of symmetrically armed (6+6 nt) hammerhead ribozymes and had the following sequences (N denotes a randomized position): GUC-specific library, (N)₆GAC(N)₄; GUA-specific library, (N)₆TAC(N)₄; AUC-specific library, (N)₆GAT(N)₄; AUA-specific library, (N)₆TAT(N)₄. Specific 13mer antisense ODNs targeting the selected ribozyme cleavage sites were also synthesized. These ODNs are identified by the cleavage position of their corresponding ribozymes, and have the following sequences: O-118, GGGAAGGACCTTG; O-238, GGCTTTGACAGGG; O-287, TCGGGGGACCGGG; O-465, ATGGCATACGGCT; O-519, AGCTCCGACTCAC; O-539, GCACCCGACGCAG; O-548, TGGCATGACGCAC; O-712, ATACCGGACGCCA.

Plasmids

The cDNA of the coding sequences of PSKH1 was previously cloned in our lab and inserted into the expression vector pcDNA3 (Invitrogen) downstream of a T7 promoter, producing the plasmid pcDNA3-PSK. A *Renilla* luciferase expression plasmid, pEF1-Luc, constructed by inserting the EF1 α promoter into the multiple cloning site of the pRL-null vector (Promega), was used as an internal control in co-transfection experiments. This plasmid was a kind gift from Professor A.-B. Kolstø.

Preparation of PSKH1-luciferase fusion constructs

The full-length coding sequence of PSKH1 cDNA was cloned in-frame with firefly luciferase cDNA into the *Bgl*II/*Nco*I sites of the pGL3 Enhancer expression vector (Promega) as a *Bam*HI-*Nco*I PCR fragment, producing the plasmid pPSK-Luc. The PSK-Luc fusion was cloned into the tetracycline response plasmid pTRE (Clontech) in two steps. A *Kpn*I-*Bam*HI fragment

of pPSK-Luc was first transferred to the EGFP-1 vector (Clontech). Subsequently, an *Eco*RI-*Bam*HI fragment from this clone was excised and cloned into the same sites of the pTRE vector, giving pTRE-PSK-Luc. Finally, PSK-Luc cDNA was also cloned into the pcDNA3 expression vector as an *Eco*RI-*Xba*I fragment for higher-level expression. This expression plasmid was termed pcDNA3-PSK-Luc.

In vitro transcription of PSKH1 RNA

For *in vitro* transcription of PSKH1 RNA with T7 RNA polymerase, the pcDNA3-PSK plasmid was linearized internally with *Eco*NI, producing a 1.03 kb transcript containing 47 nt of 5' vector-derived sequence, or downstream with *Xba*I (producing a 1.37 kb transcript). Protein and salt were removed from restriction reactions using JetQuick columns (Genomed). Purified DNA was eluted in DEPC-treated water (DEPC-H₂O), precipitated with ethanol, and resuspended in DEPC-H₂O. Run-off transcription of uniformly ³²P-labeled RNA was performed in 50 μ l reactions containing 2 μ g template DNA, 0.5 mM each of GTP, ATP, CTP and UTP (Boehringer Mannheim), 50 U RNasin (Promega), 1–2 μ l 10 μ Ci/ μ l [α -³²P]rATP or rUTP (Amersham) and 50 U T7 RNA polymerase (New England Biolabs) in RNA polymerase buffer (40 mM Tris-HCl, 6 mM MgCl₂, 2 mM spermidine, 10 mM DTT, pH 7.9). Reactions were incubated for 2 h at 37°C. RNA was then desalted by centrifugation through RNase-free G50 Sephadex QuickSpin Columns (Boehringer Mannheim) and deproteinized by phenol extraction. RNA was then precipitated with isopropanol and resuspended in RNase-free water. Alternatively, RNA was purified by LiCl-precipitation (2.5 M final concentration) following transcription. Yield and concentration of RNA were calculated from the percentage of incorporated radioactivity.

Antisense ODN and RNase H-mediated in vitro cleavage of PSKH1 RNA

RNase H-mediated cleavage assays with antisense ODN libraries were performed in 10 μ l reactions containing 40 μ M ODN library, 50 nM PSKH1 mRNA and 0.25 U RNase H (Promega) in a buffer containing 40 mM Tris-HCl pH 7.5, 10 mM MgCl₂ and 10 mM DTT. A stock dilution of RNA in buffer was heated to 95°C for 60 s and then preincubated at 37°C for 15 min before adding enzyme. An 8.0 μ l mixture of RNA, buffer and enzyme was then mixed with 2.0 μ l 200 μ M ODN library, yielding the indicated final concentrations of components. Reactions were incubated for 30 min at 37°C, and quenched on ice with 10 μ l of denaturing loading buffer (8 M urea, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were heated to 95°C for 2 min prior to loading onto a 4% denaturing polyacrylamide gel. RNA size markers were prepared by run-off *in vitro* transcription of differently linearized (*Nco*I, *Xmn*I, *Bsa*AI, *Nae*I, *Bst*UI, *Fok*I and *Dde*I) pcDNA3 templates. Samples and standards were analyzed by electrophoresis for 3–4 h at 50 W at room temperature on 40 cm sequencing gels. The gels were then transferred to Whatman 3MM paper, wrapped in plastic foil and exposed overnight in a Phosphor Screen (Molecular Dynamics) prior to analysis on a radioanalytical scanner (Storm 860, Molecular Dynamics). Fragment sizes were calculated from their measured mobilities and those of the size markers. Reactions with specific antisense ODNs (200 nM, 4-fold molar excess) were performed essentially as for the ODN libraries.

Cultured cells

HeLa cells (from ATCC) were maintained in Dulbecco's Minimal Essential Medium supplemented with 2 mM glutamine and 10% fetal calf serum (all reagents from Gibco BRL). Upon thawing, the cells were grown to near confluency for 2–3 days and passaged at least once before they were used for experiments. Cells were routinely passaged every 3–4 days.

Transient co-transfections

Cells were plated in 12-well (Costar) or 24-well (Sarstedt) plates at 30–40% confluency and transfected at an estimated 60–80% confluency the following day. Transfections were performed with 100 nM ribozyme (1.23–1.30 µg/ml), 0.40 µg/ml reporter construct (pTRE-PSK-Luc or pcDNA3-PSK-Luc) and 8 ng/ml internal control plasmid (pEF1-Rluc). Nucleic acids were complexed with lipofectamine (Gibco BRL) (a final concentration of 8.2–8.5 µg/ml lipid) at a constant 1:5 w/w ratio, following optimization with the transfection agent. Complexes were formed by mixing equal volumes of medium-diluted nucleic acids (ribozyme + DNA) and lipofectamine and incubating at room temperature for 30 min. The complexes were subsequently diluted to the final volume with serum-free medium and added to pre-washed cells (250 and 500 µl for 16 and 22 mm wells, respectively) for 5 h. The complexes were subsequently replaced with full medium.

Uptake of FITC-labeled ribozymes

Uptake of FITC-labeled ribozymes under the standard transfection conditions was determined by fluorescence spectroscopy. Lipofectamine complexes were diluted with medium and chilled on ice prior to addition to cells (in 12-well plates) preincubated (for at least 30 min) either at 37°C or on ice. Following transfection for 5 h, cells were harvested by washing three times with ice cold phosphate-buffered saline (PBS) and lysed with 1.5 ml PBS-TDS (1% Triton, 0.5% deoxycholate and 1% SDS in PBS) for at least 15 min at room temperature under vigorous shaking to achieve complete lysis. Parallels were combined and fluorescence recorded on a spectrometer (LS-5, Perkin Elmer) at excitation and emission wavelengths of 492 and 522 nm, respectively.

Luciferase activity assays

Cells were harvested 24 h after initiation of transfection and washed twice in cold PBS prior to lysis for 30 min on ice with a passive lysis buffer supplied with the Dual-Luciferase® Reporter Assay System kit (Promega). Following brief centrifugation to pellet cell debris, luciferase assays were performed in white non-transparent 96-well plates (Nunc) using a plate-reading luminometer (MicroLumat Plus, EG&G Berthold) equipped with two injectors. Dual-luciferase assays were performed on 25 µl lysate supernatant. Firefly and *Renilla* luciferase activities were recorded following the respective injections of 100 µl LAR II and 100 µl Stop&Glow reagents of the assay system according to the manufacturer's instructions. The instrumental background levels of luminescence were recorded for empty wells.

RESULTS

Selection of ribozyme target sites by MFold secondary structure prediction

We have used the MFold program (version 3.0 at <http://mfold2.wustl.edu>) (7,8) to predict the secondary structure of PSKH1 mRNA as a means of selecting GUC sites for ribozyme targeting. The MFold program generated many suboptimal secondary structures that differed minimally in energy from the optimal folding. Although these secondary structures displayed varying degrees of folding differences, certain substructures could be identified that demonstrated a considerable degree of conservation. High conservation of these substructures among alternate suboptimal structures may increase their likelihood of being 'true' structures (i.e. contained in the actual secondary structure of the mRNA *in vivo*). Therefore, in selecting GUC sites for hammerhead ribozyme targeting, we have screened 10 optimal and suboptimal secondary structures of the first 1.0 kb of the PSKH1 transcript, as transcribed from pTRE-PSK-Luc, for recurring substructures. All selected target sites were located within identical substructures in at least six out of the 10 most energetically stable structures, including the optimal structure (Table 1). For some target sites, several of the suboptimal structures diverged only slightly from the consensus in ways that did not significantly change the basic characteristics of the substructure. Due to the lack of

Table 1. Features of the predicted secondary structure of selected ribozyme target sites

	guc118	guc238	guc287	gua465	guc519	guc539	guc548	guc712
Structural frequency	6/10	6/10	6/10	6/10	6/10	7/10	7/10	10/10
Major loop	5	8	4	6	3	0	3	3
Major stem	6	8	3	4	6	17	10	11
ΔG (helix) ^a	-12.7	-20.1	-5.7	-8.8	-8.1	-11.4	-9.7	-13.1
ΔG (target) ^b	-13.2	-19.4	-6.7	-9.4	-14.4	-18.1	-15.4	-8.3
ΔG (duplex) ^c	-32.9	-36.0	-40.2	-35.2	-34.1	-36.2	-34.0	-31.7

Structural frequency denotes the number of times the target site substructure occurred unchanged in the 10 energetically most stabilized foldings of the mRNA. Major loop and stem are the longest stretches of predicted unpaired and paired target sequence bases, respectively. Free energies (ΔG) were calculated for the most stable helical region within the target sequence^a, for the target sequence as a whole^b, and for the hammerhead-RNA duplex^c, as detailed in the text. Energies are in kcal/mol.

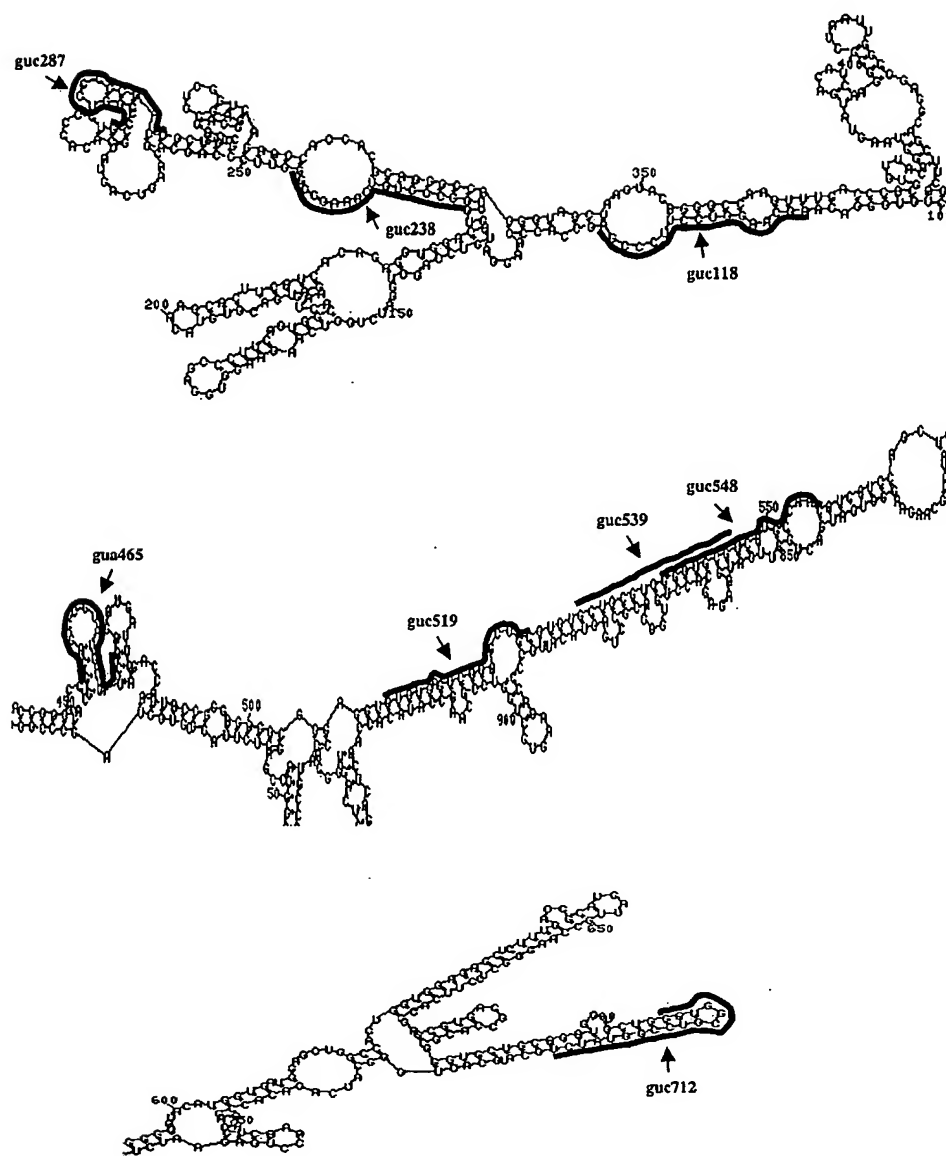


Figure 2. Excerpts of the energetically optimal secondary structure of the first 1000 bases of the pTRE-PSK-Luc transcript, as predicted by Mfold version 3.0 (7,8). Target sequences of the ribozymes are highlighted. Arrows indicate position of cleavage.

a consensus on what kind of structures make the best targets, we have chosen GUC targets (guc238, guc287, guc519 and guc712) that are predicted to be located within very different substructures (Fig. 2). The guc238 cleavage triplet is located within a large internally looped stem in which the 5' arm (stem I) of the corresponding ribozyme is complementary to the nucleotides of the loop. The guc287 target sequence, on the other hand, is predicted to fold into a short stem-loop structure. The guc519 target site is located at the end of what is essentially a very long basepaired region that is interrupted by occasional short bulges. A ribozyme was also designed targeting guc712, located at a particularly stable hairpin structure at the end of a very long, and presumably very stable, stem. This target site was included due to the high conservation of its secondary structure among suboptimal structures of the target RNA (identical in 10 out of 10 and 19 out of the 20 energetically most favorable structures). In addition to the above target sites,

the first GUC site in the RNA (guc118) was selected due to its close proximity to the translation initiation region (only ~20 nt downstream of the initiation codon). This region has often been targeted (12,33,47,48) because the RNA in this region might be relatively open due to the need for binding of the components of the translation machinery. The predicted secondary structure for the guc118 target consists of a stem with two internal loops (Fig. 2).

***In vitro* accessibility assays with semi-randomized ODN libraries and specific ODNs**

Target triplet-specific, semi-randomized ODN libraries were used to screen *in vitro* transcribed PSKH1 RNA for hybridization-accessible GUC, GUA, AUC and AUA sites. Accessible sites were identified by RNase H-mediated cleavage of the RNA at the RNA-DNA hybrids generated by hybridization of anti-sense ODNs to target RNA (19-21). Screening of a 1.03 kb

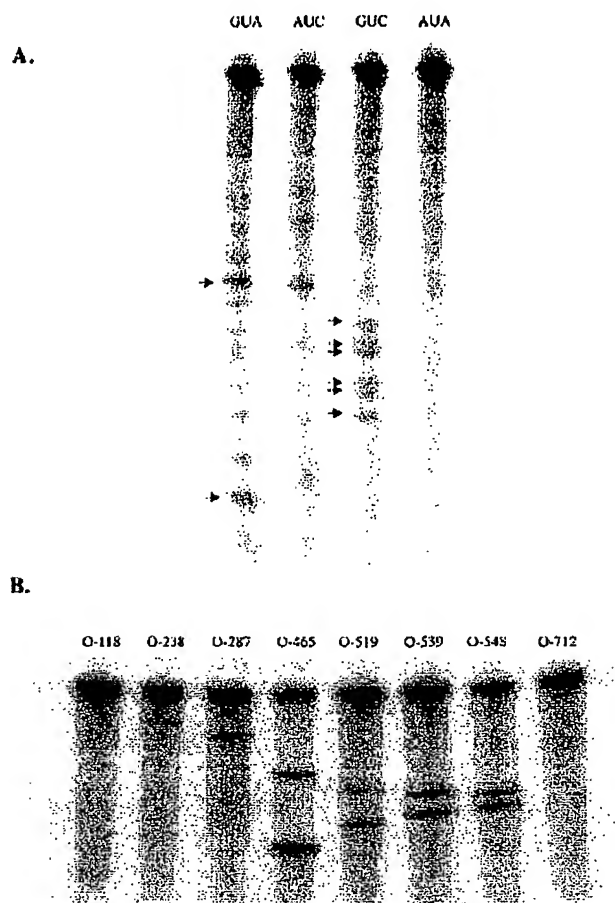


Figure 3. *In vitro* accessibility assays. (A) Separate screening of a 1.03 kb PSKH1 *in vitro* transcript with triplet-specific ODN libraries. Reactions were performed with 50 nM RNA and 40 μ M ODN library as detailed in Materials and Methods. Arrows indicate pairs of fragments resulting from cleavage at subsequently selected ribozyme target sites. (B) Cleavage of PSKH1 RNA (50 nM) with molar excess (200 nM) of specific ODNs against ribozyme target sites selected by oligo library screening (O-465, O-519, O-539, O-548) and structural considerations.

PSKH1 transcript from the pcDNA3-PSK plasmid was performed separately with each of the four ODN libraries (Fig. 3A). Since the transcript was uniformly labeled, cleavage at any position produced two cleavage products. This complicated the interpretation of the data since each pair of cleavage fragments suggested two possible sites of cleavage. Screening reactions were subsequently performed on a longer RNA transcript (1.37 kb) that differed from the first transcript in its 3'-end. This facilitated the interpretation of cleavage data, as 5' cleavage fragments were identical for the two RNAs. Screening of two differently sized transcripts increased confidence in the results. Any accessible sites that are not identified in the longer transcript are likely to be due to structural features of the specific transcript that may not be represented in the full-length mRNA. Such sites are more likely to be encountered at the ends of the transcripts since the local folding of these regions will be most strongly affected by the lack or presence of additional sequences. Consequently, any cleavage sites originating within the first or last 200 nt of the transcript were ignored in the

selection of accessible sites by library screening, as were cleavage sites that could not be detected in the longer transcript.

The screening assays demonstrated several reproducible cleavage sites that obeyed the above restrictions. Most cleavage sites were identified with the GUA and GUC libraries, while no useful sites were identified with the AUA library (Fig. 3A). Screening of PSKH1 RNA with the GUA library produced two very prominent cleavage fragments that were not produced by any of the other libraries, demonstrating their GUA triplet specificity. The sizes of these two fragments were estimated at ~615 and 405 nt. Their combined estimated size of 1020 nt was in good agreement with the transcript length (which was 1030 nt). The shorter fragment was shown to be the 5' cleavage fragment, as it was also present when screening the longer RNA transcript (data not shown). The size of this fragment was consistent with cleavage near a site corresponding to position 460 in the PSKH1 cDNA. Examination of the sequence indicated a suitable GUA triplet at position 463–465 as the only probable site of cleavage. Screening with the GUC library produced six closely spaced fragments with estimated sizes ranging from 470 to 560 nt which fitted nicely into three pairs with combined sizes of the expected length (1020–1030 nt). By a similar analysis as for the GUA library cleavage fragments, the putative cleavage sites were identified as *guc519*, *guc539* and *guc548*, respectively. One of these sites, *guc519*, had previously also been selected on the basis of MFold secondary structure predictions (see above) prior to the antisense library screening. The AUC library screening also produced several fragment pairs, but these were either weaker than their GUA and GUC counterparts, or could not be unambiguously assigned to a specific triplet. The screening assays were performed with a ratio of ODN (40 μ M) to target RNA (50 nM) of 800:1, which given the complexity ($4^{10} = 1 \times 10^6$) of the libraries corresponds to a ratio of specific ODN (40 pmol) to RNA of roughly 1:1250. This might seem too low a ratio to reasonably account for the degree of cleavage observed. However, the observed cleavage would correspond to a turnover of only ~100 molecules/h. Furthermore, the concentration of ODNs that can productively hybridize to a given target site may be significantly higher as mismatches corresponding to the ends of the ODNs are not expected to severely impair hybridization efficiency. Non-triplet-specific cleavage events (involving ODNs with mismatched or wobble-paired triplets, as well as partial hybridization of only the random-armed portions) may also conceivably occur, although they are less likely to account for the major cleavage events. Non-specific cleavage events would also be expected to result from screening with more than just one library. Examples of such cleavage events were observed, but did not include the selected cleavage sites. Nonetheless, to ensure that putative ribozyme target sites were properly identified, the provisional identifications were subsequently confirmed by cleavage of the RNA with the corresponding specific antisense ODNs. In all cases, the cleavage fragments produced by the specific ODNs co-migrated with the corresponding cleavage fragments produced by the library screenings (data not shown). These results demonstrate the utility of triplet-specific semi-randomized ODN libraries in identifying potential target sites.

In vitro cleavage assays were performed with specific ODNs against all sites targeted by ribozymes to investigate whether the sites that were selected by library screening were more

accessible than target sites selected by alternative means. The specific ODNs, like those of the semi-randomized libraries, were designed to have the same recognition sequences as a corresponding hypothetical 6+6 armed hammerhead ribozyme. The *in vitro* cleavage assays demonstrated that all the ODNs targeting sites that were selected by library screening resulted in stronger cleavage than the best among the ODNs targeting sites selected by theoretical/structural considerations (Fig. 3B). Furthermore, the strongest cleavage was achieved with the ODN (O-465) corresponding to the GUA site that was cleaved most strongly in the library screenings (Fig. 3A). Thus accessibility data obtained from library screening assays were representative for the hybridization efficiencies of the corresponding unique ODNs. Consequently, major cleavage events do not appear to be substantially influenced by the presence of non-specific ODNs in the library screening. This suggests that neither positive nor negative cooperativity of binding (49) occurs to a significant degree under our assay conditions.

Cellular uptake of ribozyme

FITC-labeled control ribozymes were used to investigate the cellular uptake of ribozyme under the optimized co-transfection conditions by spectroscopy. In order to distinguish between intracellular uptake and membrane association by spectroscopic analysis, uptake experiments were performed in parallel by incubating cells on ice as well as at 37°C. Similar amounts of MRz-FITC and ARz-FITC ribozymes were found to be associated with cells following incubation both on ice (26–28% of total added ribozyme) and at 37°C (58–59% of added ribozyme) (data not shown). The increase in cell-associated fluorescence at the higher temperature, suggests that in excess of 30% of ribozyme has been internalized. The cellular association of ribozyme following incubation on ice was shown to have a linear dependency on ribozyme concentration in the range 0–100 nM (data not shown). This is consistent with a state of equilibrium between membrane-bound and free (in medium) ribozyme. Assuming that this equilibrium is maintained upon internalization of ribozyme at the permissive temperature, the amount of internalized ribozyme is determined by the formula $I = T(\beta - \alpha)/(T - \alpha)$, where T is the total amount of ribozyme added to cells, and α and β are the measured cell-associations upon incubation on ice and at 37°C, respectively. Net uptake of ribozyme under the given transfection conditions was estimated at 42–45% (data not shown). Co-transfection optimization experiments performed at various w/w ratios (1:1 to 6:1) of lipofectamine to total nucleic acids demonstrated that uptake was only slightly influenced by lipofectamine concentration above a certain threshold (2:1 w/w ratio) that constitutes a molar excess of positive charges (from cationic liposomes) in the complexation mixture. Below this threshold (at a 1:1 w/w ratio, yielding complexes of net negative charge), uptake was severely impaired (data not shown).

Efficacy of ribozymes in co-transfection experiments in HeLa cells

The *in vivo* efficiency of the selected ribozymes was evaluated in a cell culture assay in which 100 nM of ribozyme was co-transfected with a plasmid coding for a PSKH1–luciferase fusion protein. In order to correlate for transfection variability and improve experimental reproducibility, a plasmid coding for *Renilla* luciferase (Rluc) under the control of an EF1 α -promoter

was added to the transfection mixture as an internal control. This control proved to be essential as the general expression levels of *Renilla* luciferase varied up to 2–3-fold within the same experiment for different ribozymes (data not shown). Experiments were performed as far as possible with the full complement of ribozymes of identical chemistry so that in each experiment, the efficacies of ribozymes were compared under identical conditions. In each experiment, the relative firefly luciferase activity for all ribozymes was normalized to the levels for the control ribozyme (the expression of which was set at 100%). Normalization was always performed relative to the relevant control ribozyme. The data from experiments with both methylated and allylated ribozymes are summarized in Figure 4. The most effective methylated ribozyme in co-transfection assays proved to be MRz-287, targeting *guc287*. This ribozyme reduced the level (normalized) of luciferase expressed from the reporter gene to ~23% of the control levels, i.e. the levels in cells treated with irrelevant control ribozyme. The second most efficient ribozyme was MRz-519, which resulted in ~65% inhibition of expression. The ribozymes targeting sites *guc118* and *guc548*, exhibited similar inhibitory effects, with residual expression levels just over 40%. At the other end of the spectrum we find the ribozymes targeting *guc712* and *guc539*, which result in only ~30% inhibition.

Three of the ribozymes were synthesized also in the allylated version to evaluate any systematic differences in inhibitory capacity of methylated and allylated ribozymes. The effect of the allylated ribozymes varied over a relatively wide range, resulting in residual reporter gene expression ranging from ~35% for ARz-519 to 95% for ARz-539 (Fig. 4). Comparing the data for these ribozymes with those of their methylated counterparts, we observed the same activity ranking. For both types of modification a ribozyme targeting the *guc519* cleavage site was superior to the ribozymes targeting the two other sites, *gua465* and *guc539*, of which the latter site appeared to be least amenable to cleavage. This constitutes evidence that the observed inhibitory effects of these ribozymes are sequence specific. The difference in activity of methylated and allylated versions of otherwise identical ribozymes varies with the target of the ribozymes. While the two types of ribozymes targeting *guc519* are equally active, the allylated ribozymes targeting *gua465* and *guc539* are less active than their methylated counterparts. The combined data suggest that methylation generally results in superior ribozymes and, further, that the allyl modification seems to be more detrimental to ribozyme efficacy when targeting poorly accessible sites. Reduced efficacy of allylated ribozymes is not due to differences in intracellular uptake, as the methylated and the allylated FITC-labeled ribozyme were internalized to the same level by lipofectamine-mediated transfection (see uptake experiments).

Ribozyme effects depend on ribozyme concentration and target gene promoter strength

The concentration dependence of the ribozyme effect was investigated in co-transfection experiments in which the concentration of the active anti-PSKH1 ribozyme was varied and the total concentration of ribozyme adjusted to 100 nM with control ribozyme. For dose-dependence experiments, the most active allylated (ARz-519) and methylated (MRz-287)

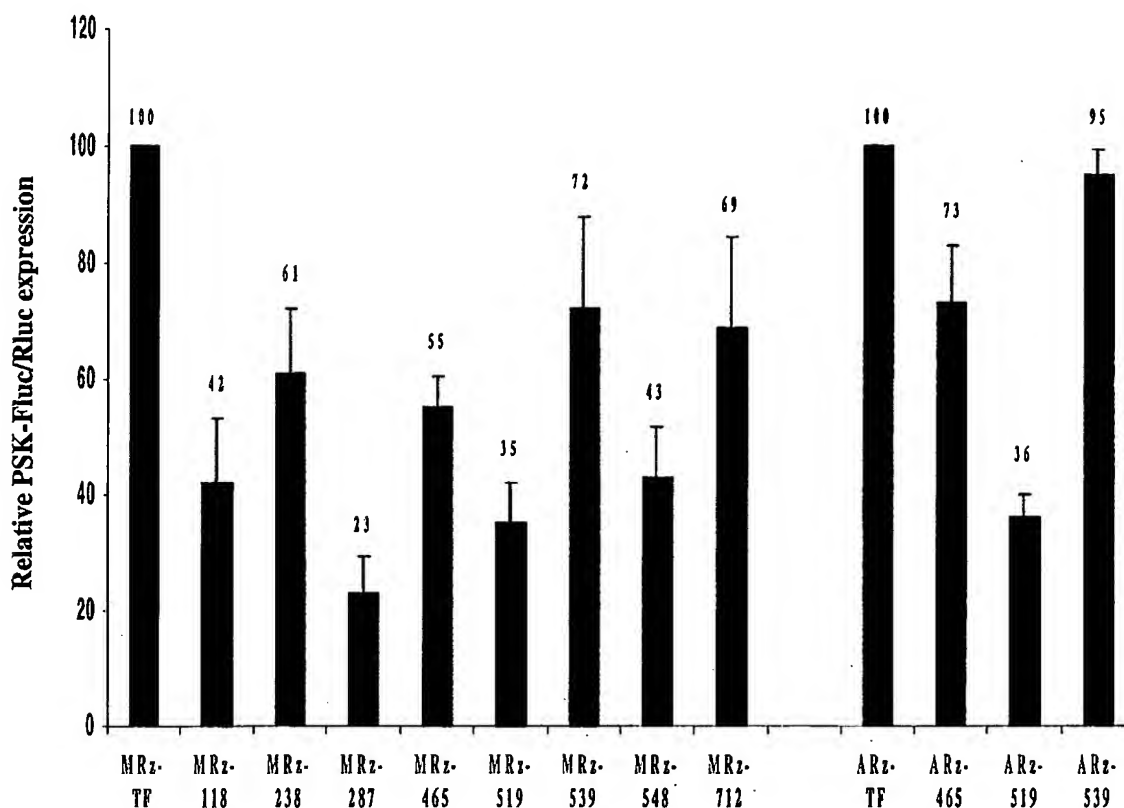


Figure 4. Lipofectamine-mediated co-transfection of HeLa cells with ribozyme, pTRE-PSK-Luc, and pRLuc as detailed in Materials and Methods. Eight methylated (MRz-) and three allylated (ARz-) PSKH1-specific ribozymes were analyzed together with their respective controls (MRz-TF and ARz-TF). PSKH1-dependent firefly luciferase expression was normalized to *Renilla* luciferase expression for each sample. Normalized expression in cells transfected with control ribozymes within each series was set at 100%. Data are averages of 4–7 independent experiments. Expression levels are indicated above the error bars (+SD) of each column.

ribozymes were selected. These experiments confirmed the dose-dependence of the ribozyme effect (Fig. 5A). We next wanted to investigate whether increasing the steady-state target gene expression would reduce the specific inhibitory effect of ribozymes. Higher expression of the reporter was achieved by placing it under the transcriptional control of the CMV promoter in the pcDNA3 expression plasmid. This resulted in 20-fold enhancement of reporter expression (data not shown) compared to the weaker tetracycline-responsive promoter. In co-transfection experiments, this increase in reporter expression was accompanied by reduced apparent efficiencies of three of the most active PSKH1 ribozymes (Fig. 5B). The average levels of inhibition of target gene expression with the weak promoter were 73, 63 and 51% for ribozymes MRz-287, MRz-519 and MRz-548, respectively. With the strong promoter, yielding the 20-fold higher expression of reporter (measured in control ribozyme-treated cells), their inhibitory activities were 60, 43 and 33%, respectively. Thus the inhibitory effect of ribozymes, as expected, was dependent on the concentrations of both ribozyme and target, but the ribozymes were able to inhibit almost completely the 20-fold increase in the target mRNA.

Comparing ribozymes selected by MFold secondary structure prediction and *in vitro* accessibility assays

Ribozymes were ranked based on their inhibitory activity in co-transfection assays, with tied ranks given to ribozymes

resulting in residual expression levels differing by <3% (Table 2). This classification can be used to assess the relative merits of the two methods of target site selection. The four ribozymes that were selected on the basis of library screening data (MRz-465, MRz-519, MRz-539, MRz-548) were compared to those selected on the basis of structural considerations (MRz-118, MRz-238, MRz-287, MRz-519, MRz-712) by Wilcoxon's rank sum test on unpaired samples (50). The rank sums for the two groups of ribozymes were 18 (average rank = 4.5) and 20 (average rank = 5.0), respectively, which are higher than the critical rank sum of 11 required for a 5% significance level. From these data, ribozyme targets selected on the basis of *in vitro* accessibility are no more susceptible to ribozyme cleavage *in vivo* than those selected by structural prediction or other considerations.

Correlation of ribozyme efficacy with the predicted target secondary structure

The two methods of target selection both resulted in ribozymes with very diverse inhibitory activities. We therefore decided to investigate whether there was any correlation between predicted structure of target sites and the *in vivo* efficacy of corresponding ribozymes. We approached this by attempting to break down the secondary structures of the target sequences in readily quantifiable parameters. In doing so we have disregarded certain factors that may influence hybridization

Table 2. Rank correlations of ribozyme efficacy with various features of the predicted target site secondary structure

Ribozyme	Efficacy	ΔG (helix)	Stem	Stem+helix	Loop	ΔG (target)	ΔG (duplex)
MRz-287	1.0	1.0	1.0	1.0	4.0	1.0	1.0
MRz-519	2.0	2.0	3.5	2.8	6.0	5.0	5.0
MRz-118	3.5 (3.0)	6.0	3.5	4.8	3.0	4.0	7.0
MRz-548	3.5 (4.0)	4.0	6.0	5.0	6.0	6.0	6.0
MRz-465	5.0	3.0	2.0	2.5	2.0	3.0	4.0
MRz-238	6.0	8.0	5.0	6.5	1.0	8.0	3.0
MRz-712	7.5 (7.0)	7.0	7.0	7.0	6.0	2.0	8.0
MRz-539	7.5 (8.0)	5.0	8.0	6.5	8.0	7.0	2.0
Pearson's correlation coefficient (r)		0.75 (0.69)	0.77 (0.80)	0.84 (0.83)	0.16	0.35	0.18
Significance level (P)		0.025 (<0.05)	<0.025	0.01	>0.05	>0.05	>0.05

Ribozymes were ranked according to inhibitory activity in co-transfection assays (lowest rank for highest activity). Target sites were ranked according to free energy of limiting helix and target sequence (lowest rank for highest energy), stem length (lowest rank for shortest stems), loop length (lowest rank for largest loops), and the free energy of the ribozyme–RNA duplex (lowest rank for lowest energy) (Table 1). Tied ranks were allotted to targets with the same length of stem and loop and to ribozymes resulting in residual expression levels differing by at most 3%. When considering stem and helix together, the average of the two individual ranks was used. Correlation coefficients have also been calculated for the case in which ribozymes were ranked without ties (numbers in parentheses). This did not significantly alter the results.

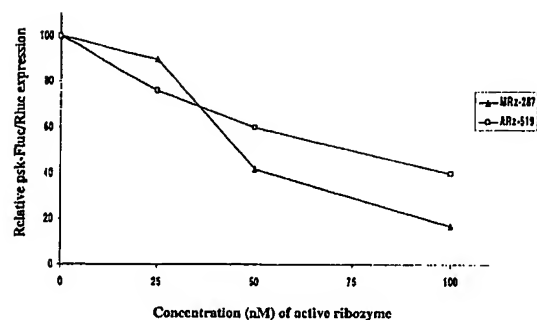
efficiency but which cannot be easily quantified. Such factors include the positioning of single-stranded stretches within the target sequence and the ability of a specific hammerhead sequence to assume its proper catalytic three-dimensional structure. The presence of so-called free ends, the positioning of single-stranded stretches at the ends of target sequences, may be expected to be correlated with enhanced antisense binding (12), while elements of strong secondary structure in the neighborhood of the target sequence proper might impede ribozyme folding. We do not imply that these factors are unimportant, merely that as they cannot be easily quantified, other more readily quantifiable parameters should be investigated for predictive value. Three potentially important parameters for target site accessibility were local free energy of folding of the target sequence (11), the size of single-stranded stretches (loops) that may function as 'hooks' for nucleation of duplex formation (51,52), and the length and stability of stems and helices. Stems and helices may need to be opened up for full hybridization of ribozyme to occur and their length and stability may therefore influence hybridization efficiency. Consequently, the target sequences of all ribozymes were decomposed into length of the major loop, stem and helix, while free energies have been calculated both for the target sequence as a whole and for its major helical region (Table 1). Target sequence free energies were calculated for the energetically optimal folding of the RNA, by adding up the energy contributions from all base-pairings, stacking interactions, bulges and various loops that are contained within the target sequences, as indicated by the MFold program. MFold was also used to determine the free energy of the most stable ('limiting') helical region and to estimate the strength of the ribozyme-target hybrid (by folding the corresponding *cis*-acting ribozyme in which catalytic and substrate strands were connected through a 5 nt loop at stem I).

Spearman's rank correlation test (53) was used to investigate the level of correlation of ribozyme efficacy ranking with various features of the predicted secondary structure of the target sequences (Table 2). No correlation was observed with the length of the major loop (Pearson's correlation coefficient $r = 0.16$) or ribozyme–substrate duplex free energy ($r = 0.18$). The correlation with target sequence free energy was weak ($r = 0.35$) and not significant. Ribozyme efficacy was, however, significantly correlated ($P < 0.025$) with both the length of the major base-paired stretch ($r = 0.75$) and the energy of the most stable helix ($r = 0.77$) within the target sequence (Table 2). Correlation was improved ($r = 0.84$) when considering stem length and helix stability together (for this analysis, the rank was taken as the average of the two individual ranks for stem and helix). Correlation coefficients were not significantly affected by ranking ribozyme activity without the use of ties (Table 2).

DISCUSSION

In this study we have analyzed DNA-armed chemically modified hammerhead ribozymes targeting eight GUC and GUA sites selected by two different methods, *in vitro* accessibility assays and MFold prediction. In a co-transfection controlled assay, the ribozymes resulted in residual luciferase reporter gene expression ranging from 23 (MRz-287) to 72% (MRz-539) (Fig. 4). This activity may in part be due to the presence of DNA arms which allow RNase H-mediated cleavage of DNA–RNA hybrids. Target gene expression normalized to the expression of a co-transfected non-target gene allowed the control of any non-target-specific sequence effects of the ribozymes. Non-specific effects are occasionally encountered in association with extended stretches of phosphorothioate (P=S) linkages (19). In an attempt to minimize such effects, no more than two consecutive P=S linkages were incorporated. Finally,

A.



B.

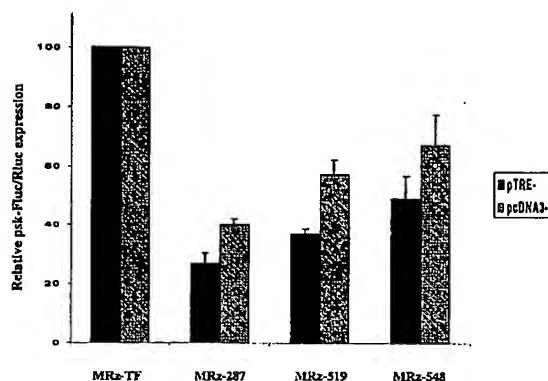


Figure 5. Dependence of ribozyme inhibitory activity on ribozyme concentration and reporter gene expression. (A) Transfection of HeLa cells with increasing concentrations of ARz-519 or MRz-287 ribozymes. Total concentration of ribozyme is adjusted to 100 nM with control ribozyme (ARz-TF or MRz-TF) and transfections performed after the standard protocol. Data for representative experiments are shown. (B) Parallel co-transfections of HeLa cells with selected ribozymes and different reporter constructs (pTRE-PSK-Luc and pcDNA3-PSK-Luc). Relative expression levels in control cells were 20-fold higher with the pcDNA3-PSK-Luc construct compared to pTRE-PSK-Luc.

ribozymes of different chemistry (with methylated RNA instead of DNA in the arms) against two of the targets (guc519 and guc548) were less efficient in inhibiting target gene expression than their DNA-armed counterparts (unpublished data). In combination, the above data make it unlikely that the observed effects are due to aspects of ribozyme sequence or chemistry unrelated to hybridization-specific activity.

The inhibitory effect of ribozymes was dependent on both the ribozyme concentration (Fig. 5A) and the steady-state expression levels of the reporter gene (Fig. 5B). A reduced inhibitory effect was observed for three selected methylated ribozymes when expressing the reporter gene from a stronger promoter resulting in 20-fold higher expression levels (Fig. 5B). However, the molar amount of target suppression when the stronger promoter was used increased to a degree almost matching the increase in the level of the exogenous target. This interesting fact suggests an enzymatic rather than a stoichiometric effect of the ribozymes on target gene expression. The enzymatic function may derive entirely from the ribozyme or include RNase H activity. It should, however, be noted that the increase in the total level of PSKH1 mRNA by expressing

the exogenous transcript from the stronger promoter depends also on the unknown level of the endogenous transcript, which is another target for the ribozyme. The level of suppression achieved with the best ribozyme in this study (77%) is similar to that which has previously been reported for the best of several unmodified ribozymes in a similar luciferase reporter gene co-transfection assay (16). Other comparable studies have reported inhibition levels in the range of 40–80% for a set of 15 ribozymes (54), 50% inhibition with a pair of variously modified ribozymes (29), and 40–55% inhibition for a pair of 2'-F-pyrimidine modified ribozymes with phosphorothioates at both ends (27). Thus our results, achieved with generally more extensively modified ribozymes, compare favorably with previous reports.

We have also attempted to evaluate the relative effects of two commonly employed types of 2'-O-modifications on the activity of ribozymes. The same rank order of ribozymes targeting three selected sites was observed for both methylated and allylated DNA-armed ribozymes (Fig. 4). Allylated ribozymes were generally less efficient in inhibiting reporter gene expression in co-transfection assays compared to their methylated counterparts. The difference in activity of methylated and allylated versions of otherwise identical ribozymes appeared to correlate with the susceptibility of the target to inhibition by ribozyme. While the ribozyme targeting the least accessible of the dually target sites (guc539) apparently was most sensitive to the type of alkylation, ribozyme species of either modification were equally efficient when targeting the most susceptible site (guc519).

Predicted secondary structures of targets selected on the basis of the MFold program were very diverse and included a short stem-loop (guc287), a stem with a large internal loop (guc238), a bulged stem (guc519), and a hairpin structure (guc712), as well as a site near the translation initiation site that was presumed to be relatively unstructured (Fig. 2). *In vitro* accessibility assays with cleavage triplet-specific ODN-libraries identified three accessible GUC sites and one GUA site (Fig. 3A). *In vitro* cleavage assays with specific ODNs against all sites targeted by ribozymes confirmed that the sites that were selected by library screening were indeed more accessible *in vitro* than sites selected by alternative means (Fig. 3B). This confirms the utility of such semi-randomized libraries for identifying the most accessible sites *in vitro*. However, ribozymes targeting sites selected on the basis of *in vitro* accessibility assays were no more efficient in inhibiting target gene expression in a co-transfection assay than ribozymes targeting sites selected by theoretical means. In fact, the ribozyme targeting the most accessible site *in vitro* (MRz-465) ranked only as the fifth most active (Table 2), while the target site of the best ribozyme (MRz-287), was relatively inaccessible *in vitro* (Fig. 3B). The lack of correlation between *in vitro* accessibility and *in vivo* efficacy data is consistent with previous observations of other researchers (55–58). A possible explanation for the poor correlation is that the target mRNA is folded differently *in vitro* than *in vivo*. The secondary structure of the folded RNA may not be the energetically most stable (fast local folding events may prevent more energetically favorable interactions between distal regions). Structural features that promote hybridization *in vivo* and *in vitro* may also differ. Furthermore, the hybridization efficiency of 13mer antisense ODNs may not be entirely representative for the hybridization

characteristics of longer (32mer) hammerhead ribozymes with varying degrees of secondary structure of their own. Finally, the generation of higher-order mRNA structures and modulation or masking of the mRNA secondary structure by RNA-binding proteins (9,10) or protein complexes (ribosomes) (59) *in vivo* may influence accessibility of target sites, although these factors would also tend to rule out the applicability of secondary structure predictions for target site selection. Recent data suggest that performing *in vitro* accessibility assays on an endogenous transcript in a protein environment (cell extracts) may improve the accuracy of the predictions (14,22).

Although *in vitro* accessibility assays proved to be of limited predictive value for the *in vivo* situation, correlative studies suggested that secondary structure predictions might have some merit. Significant correlation was found between ribozyme efficacy and the presence of short stems and energetically unstable helices within the ribozyme target sequence (Table 2). Ranking of ribozymes according to these two criteria, the relative efficacies of ribozymes were predicted nearly perfectly, the only significant discrepancy being the ribozyme targeting *gua465*. As well as being most accessible *in vitro*, this site also has a secondary structure that according to the above criteria should make it a significantly better target site than observed here. Other factors, such as the lack of single-stranded bases near the ends of the target sequence or a prohibitive environment for ribozyme folding, may explain the results. Notwithstanding this discrepancy, there is an apparently clear correlation between ribozyme efficacy and predicted target sequence secondary structure. Furthermore, the combined data from this study suggest that the previously reported correlation of ribozyme efficacy with local folding potential (11) may be incidental. Our data suggest that high target sequence free energy alone may not be sufficient for efficient ribozyme targeting. We propose that the above correlation is a consequence of the need to have some unpaired regions to facilitate fast nucleation of duplex formation (51,52), combined with short base-paired regions and helices of low stability that easily open up. Fulfillment of these criteria will in many cases result in a low local folding potential (high free energy) for the target site.

A recent study on the effect of varying RNA secondary structure on the efficiency of specific antisense ODNs concluded that target sequences located within regions designed to be unstructured were most effective, while targets within stable stem-loop structures were ineffective (13). Recent reports by Patzel and co-workers (12,14) described a theoretical approach for antisense ODN target site selection based on the prediction of large single-stranded stretches (loops) by MFold. Our data do not support a correlation of ribozyme activity with the length of loops, possibly because all our targets had shorter predicted loops than recommended in the above studies. Our hypothesis and the conclusions of the above studies are, however, not mutually exclusive. Target sites containing very large loops will have a good probability of also containing short helical regions, which we propose to be the limiting factor. In fact, applying our hypothesis of target site evaluation to three previously well characterized target sites (t351, t398, t498) (14,22) within mRNA for murine DNA methyl transferase, the same rank susceptibility of targets, in perfect accordance with actual inhibition data, is predicted by both theories.

In addition to the structure of the target site, the composition of the target sequence may also be of some importance. Sequences with a high G+C content will hybridize more efficiently with the complementary arms of their ribozymes and possibly increase the efficacy of the ribozyme. Although a general correlation between hybrid stability and ribozyme efficacy was not supported by our data (Table 2), it is worth noting that the most efficient ribozyme, MRz-287, has a substantially higher affinity for its target sequence than the other ribozymes (Table 1) due to an unusually high G+C content (13 out of 16 nt). In conclusion, the target sequence of MRz-287 represents the proposed desirable structural features for a good ribozyme/antisense target site. The target sequence consists of alternating short stretches of paired and unpaired bases, which limit stems and helical regions to no more than 3 bp. All other target sequences fold into secondary structures containing longer and more stable helical regions. One study has reported that hybridization accessibility for hammerhead ribozymes is correlated with the presence of unpaired bases near the cleavage triplet (60). Although our data do not suggest this to be a critical requirement for *in vivo* activity, the above criterion is also fulfilled for the *guc287* target site, as the longest single-stranded stretch is situated around the cleavage triplet and includes the base preceding the scissile bond (Fig. 2). All of the above mentioned features of *guc287* add up to a very effective ribozyme target site, in good agreement with its observed inhibitory capacity in HeLa cells.

In conclusion, our study indicates that there is a poor correlation between the apparent *in vivo* accessibility of a target and its accessibility in a completely cell-free *in vitro* assay as performed here. Thus such assays appear to be of limited value even for a preliminary selection of target sites. However, predictions by the MFold program suggest a correlation of certain features of the predicted secondary structures of target sequences, helical stability in particular, with ribozyme efficacy. The generality of these findings will, however, need to be investigated in an alternate test system. If these correlations should be confirmed, this would represent a significant improvement in the preliminary selection of candidate ribozymes. Ultimately, however, an empirical cell-based assay will still need to be performed to select the best of these candidates.

ACKNOWLEDGEMENTS

M.A. is a research fellow of the Norwegian Cancer Society. This work was supported by grants to H.P. from the Norwegian Cancer Society and the Norwegian Research Council.

REFERENCES

1. Amarzguioui, M. and Prydz, H. (1998) *Cell. Mol. Life Sci.*, **54**, 1175–1202.
2. Ludwig, J., Blaschke, M. and Sproat, B.S. (1998) *Nucleic Acids Res.*, **26**, 2279–2285.
3. Vaish, N.K., Heaton, P.A., Fedorova, O. and Eckstein, F. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 2158–2161.
4. Milner, N., Mir, K.U. and Southern, E.M. (1997) *Nat. Biotechnol.*, **15**, 537–541.
5. Ho, S.P., Bao, Y., Leshar, T., Malhotra, R., Ma, L.Y., Fluharty, S.J. and Sakai, R.R. (1998) *Nat. Biotechnol.*, **16**, 59–63.
6. Brede, G. (2000) EMBL accession no. AJ272212.
7. Zuker, M., Mathews, D.H. and Turner, D.H. (1999) In Barciszewski, J. and Clark, B.F.C. (eds), *Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and*

- Biotechnology*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 11–43.
8. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) *J. Mol. Biol.*, **288**, 911–940.
 9. Tsuchihashi, Z., Khosla, M. and Herschlag, D. (1993) *Science*, **267**, 99–102.
 10. Sioud, M. (1994) *J. Mol. Biol.*, **242**, 619–629.
 11. Sczakiel, G., Homann, M. and Rittner, K. (1993) *Antisense Res. Dev.*, **3**, 45–52.
 12. Patzel, V., Steidl, U., Kronenwett, R., Haas, R. and Sczakiel, G. (1999) *Nucleic Acids Res.*, **27**, 4328–4334.
 13. Vickers, T.A., Wyatt, J.R. and Freier, S.M. (2000) *Nucleic Acids Res.*, **28**, 1340–1347.
 14. Scherr, M., Rossi, J.J., Sczakiel, G. and Patzel, V. (2000) *Nucleic Acids Res.*, **28**, 2455–2461.
 15. Shimayama, T., Nishikawa, S. and Taira, K. (1993) *Nucleic Acids Res. Symp. Ser.*, **29**, 177–178.
 16. Sakamoto, N., Wu, C.H. and Wu, G.Y. (1996) *J. Clin. Invest.*, **98**, 2720–2728.
 17. Szymanski, M., Fürste, J.P., Barciszewska, M.Z., Erdmann, V.A. and Barciszewski, J. (1997) *Biochem. Mol. Biol. Int.*, **41**, 439–447.
 18. Mir, K.U. and Southern, E.M. (1999) *Nat. Biotechnol.*, **17**, 788–792.
 19. Jarvis, T.C., Wincott, F.E., Alby, L.J., McSwiggen, J.A., Beigelman, L., Gustofson, J., DiRenzo, A., Levy, K., Arthur, M., Matulic-Adamic, J., Karpeisky, A., Gonzalez, C., Woolf, T.M., Usman, N. and Stinchcomb, D.T. (1996) *J. Biol. Chem.*, **271**, 29107–29112.
 20. Ho, S.P., Britton, D.H., Stone, B.A., Behrens, D.L., Leffert, L.M., Hobbs, F.W., Miller, J.A. and Trainor, G.L. (1996) *Nucleic Acids Res.*, **24**, 1901–1907.
 21. Birikh, K.R., Berlin, Y.A., Soreq, H. and Eckstein, F. (1997) *RNA*, **3**, 429–437.
 22. Scherr, M. and Rossi, J.J. (1998) *Nucleic Acids Res.*, **26**, 5079–5085.
 23. Matveeva, O., Felden, B., Audlin, S., Gesteland, R.F. and Atkins, J.F. (1997) *Nucleic Acids Res.*, **25**, 5010–5016.
 24. Zoumadakis, M. and Tabler, M. (1995) *Nucleic Acids Res.*, **23**, 1192–1196.
 25. Lieber, A. and Strauss, M. (1995) *Mol. Cell. Biol.*, **15**, 540–551.
 26. Hertel, K.J., Pardi, A., Uhlenbeck, O.C., Koizumi, M., Ohtsuka, E., Uesugi, S., Cedergren, R., Eckstein, F., Gerlach, W.L., Hodgson, R. *et al.* (1992) *Nucleic Acids Res.*, **20**, 3252.
 27. Scherr, M., Grez, M., Ganser, A. and Engels, J.W. (1997) *J. Biol. Chem.*, **272**, 14304–14313.
 28. Parry, T.J., Cushman, C., Gallegos, A.M., Agrawal, A.B., Richardson, M., Andrews, L.E., Maloney, L., Mokler, V.R., Wincott, F.E. and Pavco, P.A. (1999) *Nucleic Acids Res.*, **27**, 2569–2577.
 29. Bramlage, B., Alefelder, S., Marshall, P. and Eckstein, F. (1999) *Nucleic Acids Res.*, **27**, 3159–3167.
 30. Tuschl, T. and Eckstein, F. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6991–6994.
 31. Beigelman, L., Karpeisky, A., Matulic-Adamic, J., Haeberli, P., Sweedler, D. and Usman, N. (1995) *Nucleic Acids Res.*, **23**, 4434–4442.
 32. Lyngstadaas, S.P., Risnes, S., Sproat, B.S., Thrane, P.S. and Prydz, H.P. (1995) *EMBO J.*, **14**, 5224–5229.
 33. Gu, J.L., Veerapanane, D., Rossi, J., Natarajan, R., Thomas, L. and Nadler, J. (1995) *Circ. Res.*, **77**, 14–20.
 34. Wang, Q., Mullah, B., Hansen, C., Asundi, J. and Robishaw, J.D. (1997) *J. Biol. Chem.*, **272**, 26040–26048.
 35. Flory, C.M., Pavco, P.A., Jarvis, T.C., Lesch, M.E., Wincott, F.E., Beigelman, L., Hunt, S.W., III and Schrier, D.J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 754–758.
 36. Sproat, B.S., Rupp, T., Menhardt, N., Keane, D. and Beijer, B. (1999) *Nucleic Acids Res.*, **27**, 1950–1955.
 37. Hendry, P., McCall, M.J., Santiago, F.S. and Jennings, P.A. (1992) *Nucleic Acids Res.*, **20**, 5737–5741.
 38. Shimayama, T., Nishikawa, F., Nishikawa, S. and Taira, K. (1993) *Nucleic Acids Res.*, **21**, 2605–2611.
 39. Shimayama, T. (1994) *Gene*, **149**, 41–46.
 40. Heidenreich, O., Xu, X., Swiderski, P., Rossi, J.J. and Nerenberg, M. (1996) *Antisense Nucleic Acid Drug Dev.*, **6**, 111–118.
 41. Sinha, N.D., Biernat, J., McManus, J. and Köster, H. (1984) *Nucleic Acids Res.*, **12**, 4539–4557.
 42. Usman, N., Ogilvie, K.K., Jiang, M.-Y. and Cedergren, R.J. (1987) *J. Am. Chem. Soc.*, **109**, 7845–7854.
 43. Green, R., Szostak, J.W., Benner, S.A., Rich, A. and Usman, N. (1991) *Nucleic Acids Res.*, **19**, 4161–4166.
 44. Polushin, N.N., Pashkova, I.N. and Efimov, V.A. (1991) *Nucleic Acids Res. Symp. Ser.*, **24**, 49.
 45. Polushin, N.N., Moroch, A.M., Chen, B. and Cohen, J.S. (1994) *Nucleic Acids Res.*, **22**, 639–645.
 46. Vinayak, R., Andrus, A. and Hampel, A. (1995) *Biomed. Pept. Proteins Nucleic Acids*, **1**, 227–230.
 47. Wang, F.S., Kobayashi, H., Liang, K.W., Holland, J.F. and Ohnuma, T. (1999) *Hum. Gene Ther.*, **10**, 1185–1195.
 48. Kobayashi, H., Takemura, Y., Wang, F.S., Oka, T. and Ohnuma, T. (1999) *Int. J. Cancer*, **81**, 944–950.
 49. Jankowsky, E. and Schwenzer, B. (1996) *Biochemistry*, **35**, 15313–15321.
 50. Wilcoxon, F. (1945) *Biomet. Bull.*, **6**, 80.
 51. Homann, M., Rittner, K. and Sczakiel, G. (1993) *J. Mol. Biol.*, **233**, 7–15.
 52. Skripkin, E., Paillart, J.C., Marquet, R., Blumenfeld, M., Ehresmann, B. and Ehresmann, C. (1996) *J. Biol. Chem.*, **271**, 28812–28817.
 53. Altman, D.G. (1991) *Practical Statistics for Medical Research*. Chapman and Hall, London, UK, pp. 293–295.
 54. Macejak, D.G., Jensen, K.L., Jamison, S.F., Domenico, K., Roberts, E.C., Chaudhary, N., von Carlowitz, I., Bellon, L., Tong, M.J., Conrad, A., Pavco, P.A. and Blatt, L.M. (2000) *Hepatology*, **31**, 769–776.
 55. L'Huillier, P.J., Davis, S.R. and Bellamy, A.R. (1992) *EMBO J.*, **11**, 4411–4418.
 56. Beck, J. and Nassal, M. (1995) *Nucleic Acids Res.*, **23**, 4954–4962.
 57. Domi, A., Beaud, G. and Favre, A. (1996) *Biochimie*, **78**, 654–662.
 58. Ramezani, A. and Joshi, S. (1996) *Antisense Nucleic Acid Drug Dev.*, **6**, 229–235.
 59. Chen, H., Ferbeyre, G. and Cedergren, R. (1997) *Nat. Biotechnol.*, **15**, 432–435.
 60. Campbell, T.B., McDonald, C.K. and Hagen, M. (1997) *Nucleic Acids Res.*, **25**, 4985–4993.

A computational framework for optimal masking in the synthesis of oligonucleotide microarrays

Simon Kasif^{1,2,*}, Zhiping Weng¹, Adnan Derti¹, Richard Beigel^{3,4} and Charles DeLisi¹

¹Center for Advanced Genomic Technology (CAGT), Bioinformatics Program and Biomedical Engineering Department, Boston University, Boston, MA, USA, ²MIT Genome Center, Whitehead Institute, MIT, Cambridge, MA, USA and ³Institute for Advanced Studies, Princeton, NJ, USA and ⁴Temple University, Philadelphia, PA, USA

Received April 3, 2002; Revised August 5, 2002; Accepted August 15, 2002

ABSTRACT

High-throughput genomic technologies are revolutionizing modern biology. In particular, DNA microarrays have become one of the most powerful tools for profiling global mRNA expression in different tissues and environmental conditions, and for detecting single nucleotide polymorphisms. The broad applicability of gene expression profiling to the biological and medical realms has generated expanding demand for mass production of microarrays, which in turn has created considerable interest in improving the cost effectiveness of microarray fabrication techniques. We have developed the computational framework for an optimal synthesis strategy for oligonucleotide microarrays. The problem was introduced by Hubbell *et al.* Here, we formalize the problem, obtain precise bounds on its complexity and devise several computational solutions.

INTRODUCTION

Oligonucleotide and cDNA microarrays can monitor mRNA expression levels for tens of thousands of genes simultaneously (1). While both types of arrays are applied to the elucidation of normal and pathological cellular mechanisms, the longer probes on cDNA arrays make them more susceptible to cross-hybridization, and oligo arrays are designed to reduce cross-hybridization and improve sensitivity (2). In addition, oligo microarrays can be used to detect polymorphisms (3) and, therefore, can greatly facilitate the research and diagnosis of genetic predisposition to diseases. Several large companies, such as Affymetrix, Corning, Motorola and Samsung, have established or are establishing the capability of manufacturing microarrays in large quantities. Thus, reductions in cost or time to manufacture can have a significant impact on biotechnology and medicine.

cDNA microarrays are produced by spotting pre-made cDNA solutions onto a glass or nylon surface via physical

contact or ink-jet deposition. While oligonucleotides can also be synthesized and then spotted, oligo microarrays are usually manufactured by *in situ* synthesis, primarily via photolithography (4), and more recently by ink-jet deposition (5). *In situ* synthesis involves the consecutive addition of A, C, G and T nucleotides to the appropriate spots on the microarray. An important advantage of photolithographic synthesis over ink-jet deposition is that in a single cycle of synthesis, a nucleotide can be added to all desired spots on the array. This is achieved via photodeprotection of the target spots on the array surface with UV light prior to the addition of the nucleotide. Meanwhile, non-target spots must be protected from the UV light using physical or virtual masks. The fabrication of physical masks is a laborious and costly process and one mask is needed for each cycle of synthesis for each variety of arrays. Singh-Gasson *et al.* (6) used a digital micro-mirror device to reflect light selectively onto the desired spots of an array, the 'virtual masking' strategy. Nonetheless, the deprotection step for each cycle lasts ~5 min and photolabile nucleosides are expensive. Therefore, decreasing the number of cycles required to synthesize a given set of sequences can reduce time and cost. Here, we address synthesis optimization, i.e., optimizing the order of nucleotide addition.

The simplest strategy for synthesizing a given set of sequences is to add A bases wherever appropriate as the first base, then C, G and T bases, repeating this process for the second base, and so on. Chee *et al.* (3) noted that if K is the length of the longest oligonucleotide to be synthesized, maximally $4K$ cycles are required. Hubbell *et al.* (7) observed that it would be possible to skip a synthesis cycle if a base is not needed by any oligonucleotides, or if the oligonucleotides that require the base can still be synthesized when that base is presented again later. In a parallel publication, Tolonen *et al.* (8) observe that synthesis could be accelerated, even for a large set of oligonucleotides, if the order of base addition is tailored to the oligonucleotide sequences. Consequently, oligonucleotides can vary in length by more than one base at the end of every synthesis cycle. This observation has motivated the development of the optimal base addition strategy described in this paper.

*To whom correspondence should be addressed at 44 Cummington Street, Boston University, Boston, MA 02215, USA. Tel: +1 617 358 1845; Fax: +1 617 353 6766; Email: kasif@bu.edu

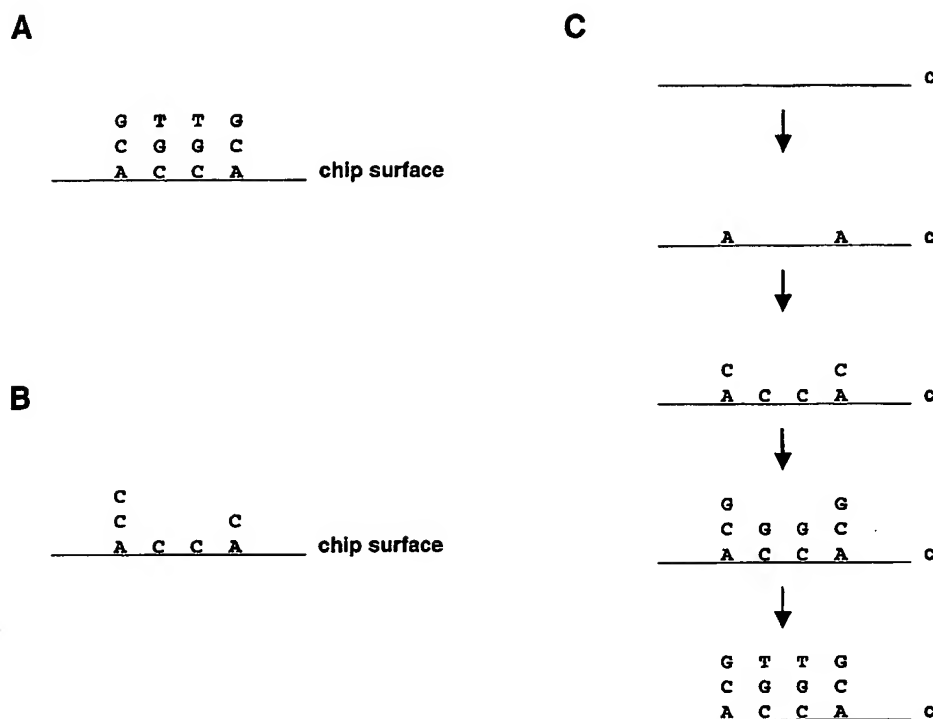


Figure 1. Formulation of synthesis strategy. (A) Four 3mer oligonucleotides. (B) Four partially constructed oligos, defining frontier $F = [3,1,1,2]$. (C) The synthesis strategy [A,C,G,T] can synthesize the array in four cycles, instead of six cycles required by the traditional approach.

A COMPUTATIONAL FRAMEWORK FOR OPTIMAL SYNTHESIS STRATEGY

We formulate the question of devising an optimal synthesis strategy in oligo microarrays as a combinatorial state space search problem. This computational formulation provides insight into the complexity of the problem and enables a range of discrete optimization and heuristic search solutions.

In this paper, we assume that the input to the optimization software is a collection of N oligo sequences of arbitrary length, which have been pre-selected in a probe selection process. For simplicity, we discuss the case of uniform length K mers, but our framework is readily applicable to the more general case. An optimal synthesis strategy involves L cycles of synthesis where, in each cycle, a single and identical nucleotide is added to all unmasked oligos. The exact spatial location of each oligo on the array is not important, as long as it can be retrieved during actual synthesis. Therefore, we assume that the input to the optimization code is a list of oligos arranged in one dimension, such as shown in Figure 1A. We define a strategy for constructing K mers in L cycles as an L long vector S , consisting of elements A, C, G and T. $S[j]$ is the nucleotide added in cycle j . For instance, the vector [A,C,A,T,G] corresponds to using A, C, A, T and G in cycles 1–5, respectively. We define the height of each partially constructed oligo as the number of nucleotides that have been added thus far by the synthesis strategy. We define a frontier F of a partially synthesized array to be an integer vector of size N where $F[i]$ is the height of the i th oligo thus far. For example, the frontier of the four oligos in Figure 1B is $F = [3,1,1,2]$.

The ‘traditional’ way to create a chip of N oligos, each of height K , is to perform $4K$ cycles of synthesis. After the addition of A, C, G and T to the appropriate spots of the array, all oligos will be one base long. We then proceed to synthesize layer two in four cycles and all oligos will be two bases long. We continue until the entire chip is synthesized in $4K$ cycles. It is easy to observe that by a slight modification of the order of base addition (8), we can expedite the above process. As a simplified example (Fig. 1C), we can synthesize an array of $K = 3$ in four cycles using a modified synthesis strategy, compared to six cycles with the ‘traditional’ approach.

In order to introduce the optimization framework for masking, we need to measure the ‘work’ that has been accomplished after several cycles of synthesis. We therefore give two definitions of ‘frontier height’: (i) the min height of a frontier constructed after L cycles is the length of the shortest oligo [in Fig. 1B, min height (L) = 1]; (ii) the sum height of a frontier is the sum of the lengths of all oligos constructed so far [in Fig. 1B, sum height (L) = $3 + 1 + 1 + 2 = 7$].

The objective optimization criterion we desire to minimize is the number of cycles required to create a frontier of min height $= K$, i.e., we seek the shortest length strategy vector that is sufficient to synthesize all oligos on the chip. It is obvious that the best possible strategy for a K mer oligo chip is of length between K and $4K$. It is easy to construct an example where the shortest strategy is of length $4K$ (Fig. 2A), although genomic sequences typically do not exhibit such extremely low complexity. In general, as the number of oligos on a chip grows, the length of the optimal strategy vector is expected to grow as well.

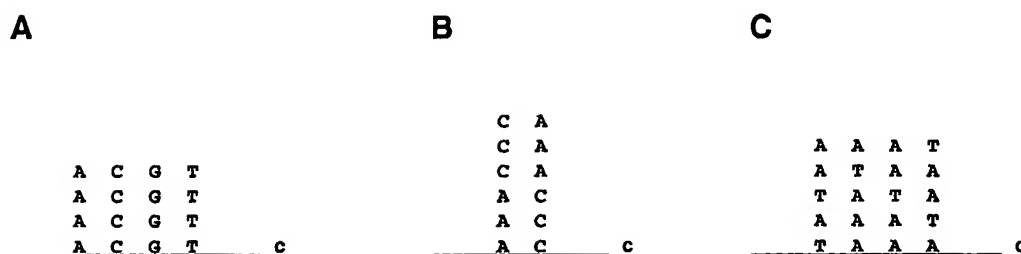


Figure 2. Example arrays that challenge synthesis strategies. (A) A worst-case scenario requiring $4K$ stages for Kmer synthesis. (B) The optimal solution for this chip requires 9 cycles whereas the greedy solution produces a 12-cycle strategy. However, the sum height-based greedy solution produces the optimal synthesis strategy. (C) For this example, sum height heuristics create an 8-cycle strategy: AATAATAA. The exhaustive search produces a 7-cycle strategy: ATAATAA.

Local Search Algorithm:

Input: A Strategy of Length M

Until No More Improvement is Observed Iterate the Following Steps:

1. For each position in I the current strategy $0 < I < M$
 - a. Try every possible NT in position I
 - b. Compute the length of the resulting strategy
2. Accept the strategy if the strategy is best over $4M$ perturbations (strategies) tried.

Figure 3. A local search algorithm.

HEURISTIC SEARCH SOLUTIONS

An obvious heuristic solution to devising the optimal synthesis strategy is a greedy search. In each synthesis cycle, we consider the four different options to extend the current layout and compute the height of the resulting frontier for each option. We choose the nucleotide that maximizes the height of the frontier, which could be either min height or sum height. The min height heuristic extends the shortest oligo, while the sum height heuristic chooses the nucleotide that will add the largest number of nucleotides to the chip.

The simple examples in Figure 2B and C show that a greedy search is not guaranteed to produce an optimal solution using either the min height or the sum height heuristic. In the remainder of this section, we consider ways to improve the greedy search. In the next section, we show that an efficient polynomial time solution is unlikely to exist for this optimization problem. We then report simulation results that describe the effectiveness of the sum height heuristics.

Look-ahead solution

A natural way to extend the greedy algorithm is to consider a look-ahead strategy. (i) Generate all possible frontiers that can be generated in L cycles. The number of strategies is 4^L . The number of frontiers might be smaller since different strategies may generate the same frontier. (ii) For each frontier, compute the height. (iii) Choose (for the first cycle) the strategy that maximizes the height after L cycles. (iv) Repeat until all oligos have been synthesized.

When $L = 4K$, this algorithm performs an exhaustive search. A rough upper bound on the running time of this algorithm is $O(4^L N)$, which makes it prohibitive for large values of L . An alternative approach would be to use a variant of best-first search such as A^* , a popular algorithm in the artificial

intelligence community. A more space-efficient alternative would be to use a branch-and-bound formulation, a standard approach in discrete optimization.

Local search solution

A local search attempts to improve a given solution by a series of local perturbations until a minimum is achieved. One obvious local search approach for our problem would be to repeatedly change a selected nucleotide in the strategy vector and accept the new strategy if it results in an improvement over the previous one. Here, we implement a variant (Fig. 3) based on steepest descent.

The steepest descent algorithm considers every possible local perturbation of a single nucleotide in all positions of the strategy vector and chooses the move that results in the greatest improvement. This algorithm terminates relatively quickly since there are at most M possible improvements to be made. Each iteration (steps 1–2) requires time $O(M)$, so the total running time is $O(M^2)$.

A simple classic variant of this algorithm is to accept a new strategy with some probability even if no improvement is observed. A Gibbs sampler is a special case of this solution when the probability of acceptance is a function of the degree of improvement, and positions for possible perturbations are selected at random rather than sequentially as described above. We describe simulation results using local search below.

COMPUTATIONAL ANALYSIS OF OPTIMAL SYNTHESIS STRATEGY

In this section, we provide a set of computational reductions that allow us to obtain a precise characterization of the complexity of the optimal synthesis strategy. We see this part as the main contribution of this paper.

Multiple sequence alignment formulation

We first observe that the optimal masking problem can be reduced to a special case of multiple sequence alignment. A precise description of multiple sequence alignment can be found in Gusfield (9) and Waterman (10). In particular, the best L cycle synthesis strategy directly corresponds to the optimal multiple alignment of the N oligos, where the costs of the alignment are defined as follows: (i) replacement cost = $+\infty$; (ii) deletion cost = $+\infty$; (iii) insertion cost = $+1$. That is, the only allowed 'editing' operation is the insertion of a gap. We first demonstrate this principle with an example. For the oligo design problem in Figure 2B, we first align the two oligos CCCAAA and AAACCC. An optimal alignment is given by

```
CCCAAA
AAACCC
```

Walking across the alignment from left to right creates the following synthesis strategy: [CCCAAACCC]. Another optimal strategy is [AAACCCAAA].

The formal proof of this equivalence is not difficult. Each strategy corresponds to a multiple alignment obtained by aligning each sequence against the strategy sequence. Therefore, the shortest strategy corresponds to the shortest global alignment.

This observation enables the application of computational solutions developed in multiple sequence analysis such as dynamic programming, Gibbs sampling and iterative refinement (9,10). A common greedy solution is based on aligning each pair, then producing a strategy based on the best aligned pair and subsequently continuing to add oligos to the alignment in the best-first manner.

Shortest super-sequence formulation

The above observation is useful for obtaining insight into the problem. By reducing our problem to a special case of multiple sequence alignment, we show that multiple alignment is 'harder', which does not preclude the possibility of an efficient solution to our specific problem. Here we show that the optimal synthesis strategy problem is exactly equivalent to the problem of computing the shortest super-sequence of a collection of strings. This two-way reduction establishes our problem to be as hard as the shortest super-sequence problem, which is known to be NP hard.

We informally define sequence X to be a super-sequence of sequence Y if every character of Y occurs in X in the same order as they occur in Y . Similarly, we define a super-sequence of a collection of sequences where the above condition has to hold for each of the sequences. For instance, AAAACCCCTTTT is a super-sequence of ACT, AAACCCT and AAAATTT. Sequence X is the shortest super-sequence of a collection of sequences if and only if its length is the shortest among all super-sequences of the collection.

Since the synthesis strategy must be a super-sequence of each of the oligos, the optimal synthesis strategy vector is equivalent to the shortest super-sequence of all oligos. The shortest super-sequence problem is known to be NP hard (11) and therefore the reduction above formally establishes the optimal synthesis strategy problem to be NP hard. More

explicitly, the problem of finding a masking strategy of length L ($L < 4K$), given a collection of N oligos of length K , is NP complete.

This is an important observation since it implies that the optimal synthesis strategy is unlikely to have efficient (sub-exponential) optimal solutions for a large number of oligos. It is easy, of course, to devise relatively efficient dynamic programming solutions when the number of oligos is constant (e.g. less than 10).

SIMULATIONS WITH RANDOM OLIGOS

We have conducted a large number of simulations to estimate the performance of several heuristic approaches to devising an optimal synthesis strategy. Here we report our results with three heuristic approaches. (i) Oblivious strategy: we simply repeat synthesizing ACGTACGT... independent of the input sequences. (ii) Max sum height heuristics: we choose the nucleotide that maximizes the sum height in the next cycle (as outlined above). (iii) Randomized local search improvement: once a solution is obtained by the above two methods, we attempt to improve the solution using local search.

Our results comparing the oblivious and max sum height heuristics are summarized in Table 1. It is clear that for random oligos there is no significant difference in performance between max height and oblivious heuristics. It is not particularly surprising for random oligos since, roughly speaking, every layer in the chip contains an approximately equal number of nucleotides of each type (A, C, G and T). Moreover, our results for 'real' oligos appear to be consistent with this performance (data not shown). Note that one of the criteria for selecting oligos aims to prevent cross-hybridization between mRNA and multiple oligos. This puts 'selective pressure' on the design to ensure that oligos are as different as possible. As the number of oligos on the chips grows, they behave more and more like random oligos.

We have interpolated the expected length of a strategy for 10 000 oligos and it appears to fit the following function well:

$$f(K) = 2.5K = 4.04\sqrt{K}$$

where K is the height of the oligos.

Figure 4 shows an essentially linear fit of the data. The graph was produced by fitting the function $f(K) = 2.5K + C\sqrt{K}$, where C is the single adjustable parameter. As a result, the fit is linear. The formal derivation that proves this expectation for max sum height is implied by the analysis in Jiang and Li (11).

Now we provide a brief motivation for the above interpolating function. When we find the shortest alignment of a single random oligo $X_1X_2...X_K$ with ACGTACGT..., X_1 aligns with the first, second, third or fourth base, X_2 aligns with the first, second, third or fourth base after X_1 , X_3 aligns with the first, second, third or fourth base after X_2 , etc. Thus the expected distance between X_i and X_{i+1} is $(1 + 2 + 3 + 4)/4 = 2.5$. For example, if $X_i = C$, then the distance between X_i and X_{i+1} is 1 if $X_{i+1} = T$, 2 if $X_{i+1} = G$, 3 if $X_{i+1} = A$ or 4 if $X_{i+1} = C$. The variance of the above possible distances is 1.25. Therefore, we expect $X_1X_2...X_K$ to require an alignment of length $2.5K$ and, by the law of large numbers, a random oligo

Table 1. Comparison of the oblivious strategy with the max sum height heuristic

<i>K</i>	<i>N</i>	Max sum height	Standard deviation	ACGT	Standard deviation	ADV
10	10 000	37.4	0.70	37.7	0.95	-3
	20 000	37.9	0.74	38.2	1.14	-3
	30 000	37.7	0.67	37.9	0.32	-2
	40 000	38.0	0.67	38.1	0.57	-1
20	10 000	68.1	0.57	67.7	1.25	4
	20 000	69.0	0.67	68.7	0.67	3
	30 000	69.7	0.67	70.7	0.82	-10
	40 000	69.7	1.06	70.1	1.20	-4
40	10 000	125.6	1.26	127.1	2.47	-15
	20 000	126.9	1.20	127.8	1.32	-9
	30 000	127.5	1.08	129.5	2.92	-20
	40 000	128.2	1.14	129.5	2.51	-13
60	10 000	181.3	1.49	183.2	2.94	-19
	20 000	183.3	1.42	184.0	1.33	-7
	30 000	183.8	1.69	184.3	1.77	-5
	40 000	184.8	1.81	185.4	2.72	-6
80	10 000	235.1	1.66	238.7	2.50	-36
	20 000	237.8	1.87	240.9	3.63	-31
	30 000	238.2	1.55	240.1	3.21	-19
	40 000	240.3	1.89	242.7	3.16	-24
100	10 000	290.6	2.01	291.8	2.10	-12
	20 000	292.6	2.17	294.1	2.33	-15
	30 000	293.3	1.49	295.8	2.57	-25
	40 000	294.6	1.17	296.6	2.27	-20

For each approach, we list the length of a strategy averaged over 10 experiments and the standard deviation. We also give the cumulative savings in nucleotides over 10 experiments (the ADV column). Note that the standard deviation is higher for the oblivious strategy (ACGT). *K* and *N* are the length and number of oligos on the microarray, respectively.

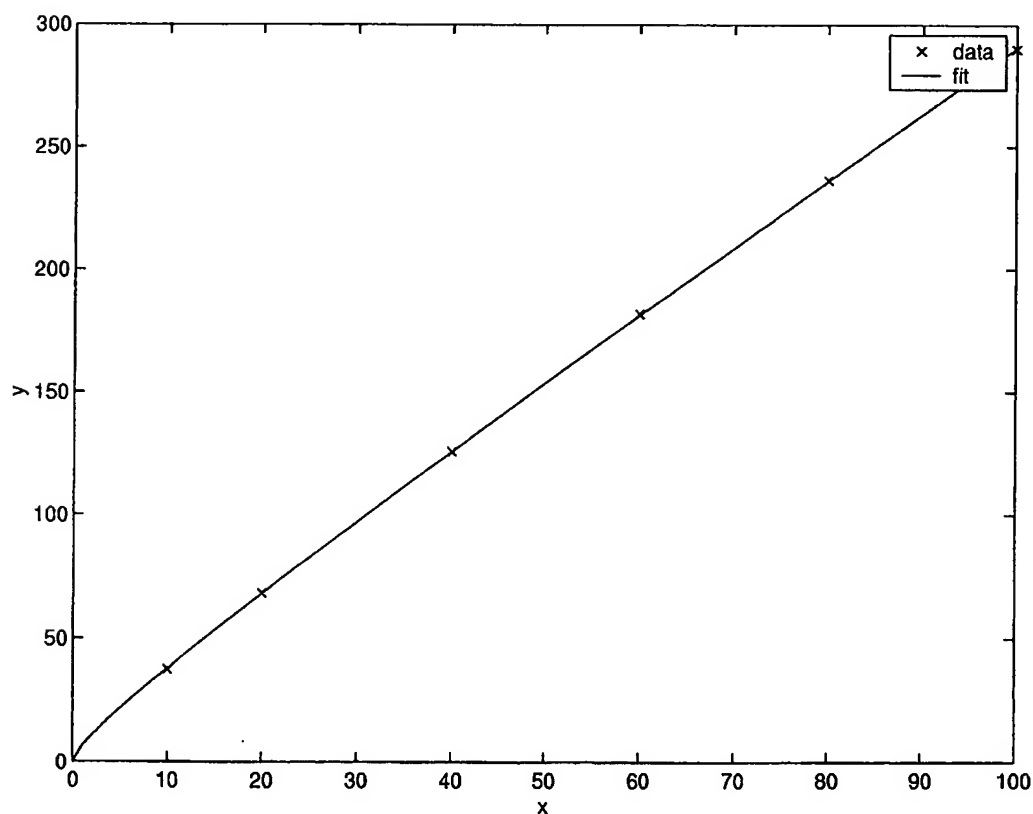


Figure 4. A linear fit of the function $f(K) = 2.5K + 4.04\sqrt{K}$ to the simulation data produced by the oblivious strategy for 1000 oligos. The scaling constant (4.04) depends on the total number of oligos.

Table 2. Results of local search improvement (LS) for both max sum height (MSH) and oblivious solutions (ACGT)

K	N	MSH	MSH-LS	ACGT	ACGT-LS
10	100	33.1	32.6	34.6	32.5
	1000	35.9	35.7	36.1	35.7
20	100	62.3	61.1	63.3	60.6
	1000	65.4	65.3	65.9	64.7
40	100	115.4	114.6	117.3	114.2

K and N are the length and number of oligos on the microarray, respectively.

$X_1X_2\dots X_K$ requires an alignment of length at most $2.5K + O(\sqrt{K})$ with high probability. In order to align N oligos, length $2.5K + O(\log N \sqrt{K})$ suffices with high probability. For a fixed number of oligos, the logarithm term is constant. Therefore, we can use the formula $2.5K + O(\sqrt{K})$.

We also used a simple local search to improve the solutions produced by the max sum height and oblivious strategies. The results are given in Table 2. We found that when the number of oligos is small (e.g. 100), the improvement was better (~3–5%). However, as the number of oligos grows (e.g. 1000), the percentage of improvement was reduced to ~1–1.5%. This is interesting, since Gibbs sampling is a close relative of local search and is a popular algorithm for multiple sequence alignment. However, in typical multiple alignments of proteins, we often align tens to hundreds of sequences. In this paper, we need to ‘align’ thousands to hundreds of thousands of oligos and it appears to have an impact on the degree of improvement obtained with this approach.

CONCLUSIONS

In this paper, we have presented a computational formalization of the optimal synthesis strategy for oligonucleotide microarrays. We have shown that the problem is computationally intractable (NP complete). We have provided several simulation results that shed light on its practical complexity. As the number of applications of oligo microarrays increases and their use in diagnostic medical applications becomes a common practice, we expect the design of DNA chips to become more sophisticated and efficient. Our main conclusion from both the theoretical and simulation analyses provided in this paper is that the problem of optimal masking appears to be computationally difficult. Moreover, the simplest possible

solution appears to work almost as well as more sophisticated approaches that include heuristic greedy approaches and local search. It would be interesting to see if more exhaustive approaches based on best-first search, branch-and-bound or Gibbs sampling methods will generate a more dramatic improvement in performance. Naturally, these results must be confirmed in the context of practically used DNA chips.

ACKNOWLEDGEMENTS

S.K. is supported in part by NSF grants IRI-9616254 and KDI-9980088. Z.W. is supported in part by NSF grants DBI-0078194 and DBI-0078194. R.B. is supported in part by NSF grants CCR-9996021 and CCR-0019019 and a grant from the State of New Jersey. A.D. was supported by the Paul and Daisy Soros Fellowship for New Americans.

REFERENCES

1. Schena, M., Shalon, D., Davis, R.W. and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
2. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. *et al.* (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **14**, 1675–1680.
3. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S. and Fodor, S.P. (1996) Accessing genetic information with high-density DNA arrays. *Science*, **274**, 610–614.
4. Lipshutz, R.J., Fodor, S.P., Gingeras, T.R. and Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21** (suppl. 1), 20–24.
5. Blanchard, A. (1998) Synthetic DNA arrays. *Genet. Eng.*, **20**, 111–123.
6. Singh-Gasson, S., Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R. and Cerrina, F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nat. Biotechnol.*, **17**, 974–978.
7. Hubbell, E.A., Morris, M.S. and Winkler, J.L. (1999) Computer-aided engineering system for design of sequence arrays and lithographic masks. US Patent no. 5 856 101.
8. Tolonen, A.C., Albeanu, D.F., Corbett, J.F., Handley, H., Henson, C. and Malik, P. (2002) Optimized *in situ* construction of oligomers on an array surface. *Nucleic Acids Res.*, **30**, e107.
9. Gusfield, D. (1997) *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, New York, NY.
10. Waterman, M.S. (1995) *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman and Hall, New York, NY.
11. Jiang, T. and Li, M. (1997) On the approximation of shortest common supersequences and longest common subsequences. *SIAM J. Comput.*, **24**, 1122–1139.

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference 13499 PCT	FOR FURTHER ACTION	see Notification of Transmittal of International Search Report (Form PCT/ISA/220) as well as, where applicable, item 5 below.
International application No. PCT/US03/36787	International filing date (day/month/year) 14 November 2003 (14.11.2003)	(Earliest) Priority Date (day/month/year) 14 November 2002 (14.11.2002)
Applicant DHARMACON INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.



It is also accompanied by a copy of each prior art document cited in this report.

1. Basis of the Report

- a. With regard to the language, the international search was carried out on the basis of the international application in the language in which it was filed, unless otherwise indicated under this item.



the international search was carried out on the basis of a translation of the international application furnished to this Authority (Rule 23.1(b)).

- b. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of the sequence listing:



contained in the international application in written form.



filed together with the international application in computer readable form.



furnished subsequently to this Authority in written form.



furnished subsequently to this Authority in computer readable form.



the statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.



the statement that the information recorded in computer readable form is identical to the written sequence listing has been furnished.

2. ☐ Certain claims were found unsearchable (See Box I).

3. ☐ Unity of invention is lacking (See Box II).

4. With regard to the title,



the text is approved as submitted by the applicant.



the text has been established by this Authority to read as follows:

5. With regard to the abstract,



the text is approved as submitted by the applicant.



the text has been established, according to Rule 38.2(b), by this Authority as it appears in Box III. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. The figure of the drawings to be published with the abstract is Figure No. 1



as suggested by the applicant.



because the applicant failed to suggest a figure.



because this figure better characterizes the invention.



None of the figures

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/36787

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/00

US CL : 435/6; 536/24.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHEDMinimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6; 536/24.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002/0150945 A1 (FINNEY et al) 17 October 2002 (17.10.2002), see page 27-28.	1-19
A	KASIF et al. A computational framework for optimal masking in the synthesis of oligonucleotide microarrays. Nucleic Acids Research. 2002, Vol. 30, No. 20, full text version of article.	1-19
A	AMARZGUIOUI et al. Secondary structure prediction and in vitro accessibility of mRNA as tools in the selection of target sites for ribozymes. Nuclei acids Research, 2000, Vol. 28, No. 21, pages 4113-4124, full text version.	1-19

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

19 January 2005 (19.01.2005)

Date of mailing of the international search report

25 FEB 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

Janet L. Epps-Ford, Ph.D.

Telephone No. 571-272-0547

Continuation of B. FIELDS SEARCHED Item 3:

CAplus, Medline, Biosis, USPatfull, Derwent, JPO, EPO

search terms: (SIRNA OR RNAI OR DSRNA) and (OPTIMIZATION OR OPTIMIZE OR OPTIMAL) and algorithm

ITEM 8

PATENT

13499US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Khvorova *et al.* Examiner: To be assigned
Serial No.: 10/714,333 Group Art Unit: 1646
Filed: November 14, 2003
For: *Functional and Hyperfunctional siRNA*
Customer No.: 23719 Kalow & Springut LLP
488 Madison Avenue, 19th Floor
New York, New York 10022

January 31, 2005

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

INFORMATION DISCLOSURE STATEMENT

Commissioner for Patents:

In accordance with 37 C.F.R. §§ 1.56 and 1.97 through 1.98, Applicants wish to make known to the Patent and Trademark Office the reference set forth on the attached form PTO-1449 (a copy of the cited reference is enclosed). As to the reference supplied, Applicants do not admit that it is "prior art" under 35 U.S.C. §§ 102 or 103, and specifically reserve the right to traverse or to antedate any such reference, as by a showing under 37 C.F.R. § 1.131 or other method. Although the aforesaid reference is made known to the Patent and Trademark Office in compliance with Applicants' duty to disclose all information of which they are aware and believe relevant to the examination of the above-identified application, Applicants believe that their invention is patentable.

Certificate of Mailing Under 37 C.F.R. 1.8

I hereby certify that this correspondence is being deposited on the date shown below with the United States Postal Service as first class mail with sufficient postage in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

K Padilla
(Signature)

K Padilla
(Printed Name of Person Signing
Certificate)

1/31/05
(Date)

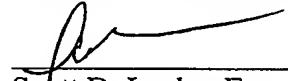
Applicants: Khvorova *et al.*
Serial No.: 10/714,333
Filed: November 14, 2003
Information Disclosure Statement
Page 2

I. FOREIGN PATENT DOCUMENTS

<u>Publication No.</u>	<u>Country</u>	<u>Publication Date</u>
WO 030646625	WIPO	7 August 2003

Because no action has been taken on the merits, Applicants submit that no fee is due at this time. However, if a fee is deemed necessary, please charge Deposit Account No. 11-0171.

Respectfully submitted,



Scott D. Locke, Esq.
Registration No.: 44,877
Attorney for Applicant

Kalow & Springut LLP
(212) 813-1600

INFORMATION DISCLOSURE CITATION*(Use several sheets if necessary)*

Docket Number (Optional)

13499 US

Application Number

10/714,333

Applicant(s)

Khvorova et al.

Filing Date

November 14, 2003

Group Art Unit

1646

U.S. PATENT DOCUMENTS

*EXAMINER INITIAL	REF	DOCUMENT NUMBER	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE

U.S. PATENT APPLICATION PUBLICATIONS

*EXAMINER INITIAL	REF	DOCUMENT NUMBER	DATE	NAME	CLASS	SUBCLASS	FILING DATE IF APPROPRIATE

FOREIGN PATENT DOCUMENTS

	REF	DOCUMENT NUMBER	DATE	COUNTRY	CLASS	SUBCLASS	Translation	
							YES	NO
		WO 03/064625	07/08/2003	WIPO				

OTHER DOCUMENTS*(Including Author, Title, Date, Pertinent Pages, Etc.)*

EXAMINER

DATE CONSIDERED

EXAMINER: Initial if citation considered, whether or not citation is in conformance with MPEP Section 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 August 2003 (07.08.2003)

PCT

(10) International Publication Number
WO 03/064625 A2

- (51) International Patent Classification⁷: **C12N**
- (21) International Application Number: PCT/US03/03208
- (22) International Filing Date: 3 February 2003 (03.02.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|------------|-------------------------------|----|
| 60/353,203 | 1 February 2002 (01.02.2002) | US |
| 60/353,381 | 1 February 2002 (01.02.2002) | US |
| 60/436,238 | 23 December 2002 (23.12.2002) | US |
| 60/438,608 | 7 January 2003 (07.01.2003) | US |

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant: SEQUITUR, INC. [US/US]; 14 Tech Circle, Natick, MA 01760 (US).

Published:

without international search report and to be republished upon receipt of that report

(72) Inventors: WOOLF, Tod, M.; 17 Pheasant Avenue, Sudbury, MA 01776 (US). TAYLOR, Margaret, F.; 232 Hayden Rowe Street, Hopkinton, MA 01748 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(74) Agents: DECONTI, Giulio, A., Jr. et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).



WO 03/064625 A2

(54) Title: OLIGONUCLEOTIDE COMPOSITIONS WITH ENHANCED EFFICIENCY

(57) Abstract: The oligonucleotide compositions of the present invention make use of combinations of oligonucleotides. In one aspect, the invention features an oligonucleotide composition including at least 2 different oligonucleotides targeted to a target gene. This invention also provides methods of inhibiting protein synthesis in a cell and methods of identifying oligonucleotide compositions that inhibit synthesis of a protein in a cell.

OLIGONUCLEOTIDE COMPOSITIONS WITH ENHANCED EFFICIENCY

5

Related Applications

This application claims the priority of U.S. provisional patent application no. 60/353,381, filed on February 1, 2002. This application also claims the priority of U.S. provisional patent application no. 60/353,203, filed on February 1, 2002, application no. 60/436,238, filed December 23, 2002, and application no. 60/438,608, filed January 7, 2003. The entire contents of the aforementioned applications are hereby expressly incorporated herein by reference.

15 Background of the Invention

Antisense and double-stranded RNA oligonucleotides are promising therapeutic agents and useful research tools for elucidating gene function. However, it is often difficult to achieve efficient inhibition of protein synthesis using such compositions.

In order to maximize their therapeutic activity, it would be of great benefit to improve upon the prior art antisense and double-stranded RNA oligonucleotides by enhancing the efficiency with which they inhibit protein synthesis.

Summary of the Invention

The instant invention is based, at least in part, on the discovery of antisense and double-stranded oligonucleotide compositions that provide improved inhibition of gene expression. In particular, the oligonucleotide compositions of the present invention make use of combinations of antisense or double-stranded oligonucleotides.

In one aspect, the invention pertains to an oligonucleotide composition comprising at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched.

In one embodiment, the oligonucleotides are antisense oligonucleotides.

In another embodiment, the oligonucleotides are double-stranded RNA oligonucleotides.

In one embodiment, the oligonucleotide compositions bind to their target nucleotide sequence with a T_m of at least about 60°C.

5 In one embodiment, the oligonucleotides have a GC content of at least about 20%.

In one embodiment, the composition comprises at least about 4 antisense oligonucleotides targeting at least four different nucleic acid sequences. In another embodiment, the composition comprises at least about 5 oligonucleotides targeting at least five different nucleic acid sequences. In still another embodiment, the composition comprises
10 at least about 6 oligonucleotides targeting at least six different nucleic acid sequences.

In one embodiment, the oligonucleotides are at least about 25 nucleomonomers in length. In another embodiment, the oligonucleotides are greater than about 25 nucleomonomers in length.

In one embodiment, at least one of the antisense oligonucleotides is complementary in
15 sequence to its target nucleotide sequence. In another embodiment, the antisense oligonucleotides activate RNase H.

In one embodiment, at least one of the oligonucleotides comprise at least one modified internucleoside linkage.

In another embodiment, at least one of the oligonucleotides comprise at least one
20 modified sugar moiety.

In one embodiment, the composition further comprises a pharmaceutically acceptable carrier.

In one embodiment, the oligonucleotide composition achieves a level of inhibition of protein synthesis the same as or higher than the level of inhibition achieved by the most
25 effective individual oligonucleotide of the composition.

In one embodiment, the individual oligonucleotides are not separately tested for their ability to inhibit protein synthesis prior to their incorporation into the composition. In this respect, the present invention represents a substantial and unrecognized improvement over the state of the art.

30 In one embodiment, the oligonucleotide composition results in greater than about 80% inhibition of protein synthesis.

In another aspect, the invention pertains to a method of inhibiting protein synthesis in a cell comprising contacting the cell with at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched, to thereby inhibit protein synthesis.

In one embodiment, the oligonucleotides are antisense oligonucleotides. In another embodiment, the oligonucleotides are double-stranded RNA oligonucleotides.

In one embodiment, the method is performed in a high-throughput format.

In still another aspect, the invention pertains to a method of identifying function of a gene encoding a protein comprising: contacting the cell with at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched, and assaying for a change in a detectable phenotype in the cell resulting from the inhibition of protein expression, to thereby determine the function of a gene.

The relative amounts of these different oligonucleotides may optionally be different. That is, the three or more different oligonucleotides may be present in equimolar concentrations, or non-equimolar concentrations.

In one embodiment, the oligonucleotides are antisense oligonucleotides. In another embodiment, the oligonucleotides are double-stranded RNA oligonucleotides.

In one embodiment, the method is performed in a high-throughput format.

In another aspect, the invention pertains to a method of making the oligonucleotide composition, comprising: combining at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched, and wherein the individual oligonucleotides are not separately tested for their ability to inhibit protein synthesis prior to their incorporation into the composition.

In one embodiment, the oligonucleotides are antisense oligonucleotides. In another embodiment, the oligonucleotides are double-stranded RNA oligonucleotides.

In another aspect, the invention pertains to an oligonucleotide composition comprising at least 3 different double-stranded RNA oligonucleotides targeted to at least three different nucleotide sequences within a target gene.

In still another aspect, the invention pertains to a method of inhibiting protein synthesis in a cell comprising contacting the cell (or cell lysate) with at least 3 different double-stranded RNA oligonucleotides targeted to at least three different nucleotide sequences within a target gene.

5 In yet another aspect, the invention pertains to a method of identifying function of a gene encoding a protein comprising: contacting the cell with at least 3 different double-stranded RNA oligonucleotides targeted to at least three different nucleotide sequences within a target gene and assaying for a change in a detectable phenotype in the cell resulting from the inhibition of protein expression, to thereby determine the function of a gene.

10 In another aspect, the invention pertains to a method of making an oligonucleotide composition comprising combining at least 3 different double-stranded RNA oligonucleotides targeted to at least three different nucleotide sequences within a target gene wherein, the individual oligonucleotides are not separately tested for their ability to inhibit protein synthesis prior to their incorporation into the composition.

15

Drawings

Figure 1 shows a summary of the results of about 30 antisense inhibition experiments against about thirty different genes in cell culture. Oligonucleotide compositions comprising mixtures of oligonucleotides (with the worst 10% of target genes removed) are compared
20 with the best individual oligonucleotides and data for all individual oligonucleotides in the percent inhibition observed.

Figure 2 shows ultramer data for a mixture of siRNA complexes targeting p53.

Figure 3 shows ultramer data for a mixture of siRNA complexes targeting GTP20.

Figure 4 shows ultramer data for a mixture of siRNA complexes targeting Cbfa-1.

25 Figure 5 shows ultramer data for a mixture of siRNA complexes targeting PTP mu.

Figure 6 shows ultramer data for a mixture of siRNA complexes targeting PTP-PEST.

Figure 7 shows ultramer data for a mixture of siRNA complexes targeting PTP eta.

Detailed Description of the Invention

Although inhibition of protein synthesis could be achieved with certain antisense and double-stranded RNA oligonucleotides of the prior art, multiple transfections were required to identify effective oligonucleotides. The instant invention advances the prior art, *inter alia*,
5 by providing oligonucleotide compositions that enhance the efficiency with which protein synthesis is inhibited and methods of making and using these improved oligonucleotide compositions.

Methods of stabilizing oligonucleotides, particularly antisense oligonucleotides, by formation of a duplex with a complementary oligonucleotide, are disclosed in co-pending
10 application no. U.S. _____, filed on the same day as the present application, bearing attorney docket number "SRI-020," and entitled "Double-Stranded Oligonucleotides." This application and all of its teachings is hereby expressly incorporated herein by reference in its entirety.

Antisense and Double-stranded RNA Oligonucleotide Compositions

Antisense or double-stranded RNA oligonucleotides for incorporation into compositions of the invention inhibit the synthesis of a target protein, which is encoded by a target gene. The target gene can be endogenous or exogenous (*e.g.*, introduced into a cell by a virus or using recombinant DNA technology) to a cell. As used herein, the term "target
20 gene" includes polynucleotides comprising a region that encodes a polypeptide or polynucleotide region that regulates replication, transcription, translation, or other process important in expression of the target protein or a polynucleotide comprising a region that encodes the target polypeptide and a region that regulates expression of the target polypeptide. Accordingly, the term "target gene" as used herein may refer to, for example, an
25 mRNA molecule produced by transcription a gene of interest. Furthermore, the term "correspond," as in "an oligomer corresponds to a target gene sequence," means that the two sequences are complementary or homologous or bear such other biologically rational relationship to each other (*e.g.*, based on the sequence of nucleomonomers and their base-pairing properties).

The "target gene" to which an RNA molecule of the invention is directed may be associated with a pathological condition. For example, the gene may be a pathogen-associated gene, *e.g.*, a viral gene, a tumor-associated gene, or an autoimmune disease-associated gene. The target gene may also be a heterologous gene expressed in a recombinant cell or a genetically altered organism. By determining or modulating (*e.g.*, inhibiting) the
35 function of such a gene, valuable information and therapeutic benefits in medicine, veterinary medicine, and biology may be obtained.

The term "antisense" refers to a nucleotide sequence that is inverted relative to its normal orientation for transcription and so expresses an RNA transcript that is complementary to a target gene mRNA molecule expressed within the host cell (*e.g.*, it can hybridize to the target gene mRNA molecule through Watson-Crick base pairing). An antisense strand may be constructed in a number of different ways, provided that it is capable of interfering with the expression of a target gene. For example, the antisense strand can be constructed by inverting the coding region (or a portion thereof) of the target gene relative to its normal orientation for transcription to allow the transcription of its complement, (*e.g.*, RNAs encoded by the antisense and sense gene may be complementary). Furthermore, the antisense oligonucleotide strand need not have the same intron or exon pattern as the target gene, and noncoding segments of the target gene may be equally effective in achieving antisense suppression of target gene expression as coding segments.

The term "oligonucleotide" includes two or more nucleomonomers covalently coupled to each other by linkages or substitute linkages. An oligonucleotide may comprise, for example, between a few (*e.g.*, 7, 10, 12, 15) or a few hundred (*e.g.*, 100, 200, 300, or 400) nucleomonomers. For example, an oligonucleotide of the invention preferably comprises between about 10 and about 50 nucleomonomers, between about 15 and about 40, or between about 20 and about 30 nucleomonomers. In one embodiment, an oligonucleotide comprises about 25 nucleomonomers. In another embodiment, an oligonucleotide comprises greater than about 25 nucleomonomers.

Oligonucleotides may comprise, for example, oligonucleotides, oligonucleosides, polydeoxyribonucleotides (containing 2'-deoxy-D-ribose) or modified forms thereof, *e.g.*, DNA, polyribonucleotides (containing D-ribose or modified forms or analogs thereof), RNA, or any other type of polynucleotide which is an N-glycoside or C-glycoside of a purine or pyrimidine base, or modified purine or pyrimidine base. The term oligonucleotide includes compositions in which adjacent nucleomonomers are linked via phosphorothioate, amide or other linkages (*e.g.*, Neilsen, P.E., *et al.* 1991. *Science*. 254:1497). Generally, the term "linkage" refers to any physical connection, preferably covalent coupling, between two or more nucleic acid components, *e.g.*, catalyzed by an enzyme such as a ligase.

The term "oligonucleotide" includes any structure that serves as a scaffold or support for the bases of the oligonucleotide, where the scaffold permits binding to the target nucleic acid molecule in a sequence-dependent manner.

An "overhang" is a relatively short single-stranded nucleotide sequence on the 5'- or 3'-hydroxyl end of a double-stranded oligonucleotide molecule (also referred to as an "extension," "protruding end," or "sticky end").

Oligonucleotides of the invention are isolated. The term "isolated" includes nucleic acid molecules which are synthesized (*e.g.*, chemically, enzymatically, or recombinantly) or are naturally occurring but separated from other nucleic acid molecules which are present in a natural source of the nucleic acid. Preferably, a naturally occurring "isolated" nucleic acid molecule is free of sequences which naturally flank the nucleic acid molecule (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid molecule) in a nucleic acid molecule in an organism from which the nucleic acid molecule is derived.

The term "nucleomonomer" includes bases covalently linked to a second moiety. Nucleomonomers include, for example, nucleosides and nucleotides. Nucleomonomers can be linked to form oligonucleotides that bind to target nucleic acid sequences in a sequence specific manner. The term "second moiety" as used herein includes substituted and unsubstituted cycloalkyl moieties, *e.g.*, cyclohexyl or cyclopentyl moieties, and substituted and unsubstituted heterocyclic moieties, *e.g.*, 6-member morpholino moieties or, preferably, sugar moieties.

Sugar moieties include natural, unmodified sugars, *e.g.*, monosaccharides (such as pentoses, *e.g.*, ribose), modified sugars and sugar analogs. Possible modifications of nucleomonomers include, for example, replacement of one or more of the hydroxyl groups with a halogen, a heteroatom, an aliphatic group, or the functionalization of the group as an ether, an amine, a thiol, or the like. For example, modified sugars include D-ribose, 2'-O-alkyl (including 2'-O-methyl and 2'-O-ethyl), *i.e.*, 2'-alkoxy, 2'-amino, 2'-S-alkyl, 2'-halo (including 2'-fluoro), 2'-methoxyethoxy, 2'-allyloxy ($-\text{OCH}_2\text{CH}=\text{CH}_2$), 2'-propargyl, 2'-propyl, ethynyl, ethenyl, propenyl, and cyano and the like. In one embodiment, the sugar moiety can be a hexose and incorporated into an oligonucleotide as described (Augustyns, K., *et al.*, *Nucl. Acids. Res.* 1992. 18:4711). Exemplary nucleomonomers can be found, *e.g.*, in U.S. Patent 5,849,902.

As used herein, the term "nucleotide" includes any monomeric unit of DNA or RNA containing a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is usually linked to the sugar moiety via the glycosidic carbon (at the 1' carbon of pentose) and that combination of base and sugar is called a "nucleoside." The base characterizes the nucleotide with the four customary bases of DNA being adenine (A), guanine (G), cytosine (C) and thymine (T). Inosine (I) is an example of a synthetic base that can be used to substitute for any of the four, naturally-occurring bases (A, C, G or T). The four RNA bases are A, G, C, and uracil (U). Accordingly, an oligonucleotide may be a nucleotide sequence comprising a linear array of nucleotides connected by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses. Other modified nucleosides/nucleotides are described herein and may also be used in the oligonucleotides of the invention.

One particularly useful group of modified nucleomonomers are 2'-O-methyl nucleotides, especially when the 2'-O-methyl nucleotides are used as nucleomonomers in the ends of the oligomers. Such 2'-O-methyl nucleotides may be referred to as "methylated," and the corresponding nucleotides may be made from unmethylated nucleotides followed by
5 alkylation or directly from methylated nucleotide reagents. Modified nucleomonomers may be used in combination with unmodified nucleomonomers. For example, an oligonucleotide of the invention may contain both methylated and unmethylated nucleomonomers.

Some exemplary modified nucleomonomers include sugar-or backbone-modified ribonucleotides. Modified ribonucleotides may contain a nonnaturally occurring base
10 (instead of a naturally occurring base) such as uridines or cytidines modified at the 5-position, *e.g.*, 5-(2-amino)propyl uridine and 5-bromo uridine; adenosines and guanosines modified at the 8-position, *e.g.*, 8-bromo guanosine; deaza nucleotides, *e.g.*, 7-deaza-adenosine; and N-alkylated nucleotides, *e.g.*, N6-methyl adenosine. Also, sugar-modified ribonucleotides may have the 2'-OH group replaced by a H, alkoxy (or OR), R or alkyl, halogen, SH, SR, amino
15 (such as NH₂, NHR, NR₂), or CN group, wherein R is lower alkyl, alkenyl, or alkynyl.

Modified ribonucleotides may also have the phosphoester group connecting to adjacent ribonucleotides replaced by a modified group, *e.g.*, of phosphothioate group. More generally, the various nucleotide modifications may be combined.

The term "alkyl" includes saturated aliphatic groups, including straight-chain alkyl
20 groups (*e.g.*, methyl, ethyl, propyl, butyl, pentyl, hexyl, heptyl, octyl, nonyl, decyl, *etc.*), branched-chain alkyl groups (isopropyl, tert-butyl, isobutyl, *etc.*), cycloalkyl (alicyclic) groups (cyclopropyl, cyclopentyl, cyclohexyl, cycloheptyl, cyclooctyl), alkyl substituted cycloalkyl groups, and cycloalkyl substituted alkyl groups. In certain embodiments, a straight chain or branched chain alkyl has 6 or fewer carbon atoms in its backbone (*e.g.*, C₁-
25 C₆ for straight chain, C₃-C₆ for branched chain), and more preferably 4 or fewer. Likewise, preferred cycloalkyls have from 3-8 carbon atoms in their ring structure, and more preferably have 5 or 6 carbons in the ring structure. The term C₁-C₆ includes alkyl groups containing 1 to 6 carbon atoms.

Moreover, unless otherwise specified, the term alkyl includes both “unsubstituted alkyls” and “substituted alkyls,” the latter of which refers to alkyl moieties having substituents replacing a hydrogen on one or more carbons of the hydrocarbon backbone. Such substituents can include, for example, alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxy carbonyloxy, carboxylate, 5 alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxy, phosphate, phosphonato, phosphinato, cyano, amino (including alkyl amino, dialkylamino, arylamino, diarylamino, and alkylaryl amino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, 10 alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety. Cycloalkyls can be further substituted, *e.g.*, with the substituents described above. An “alkylaryl” or an “arylalkyl” moiety is an alkyl substituted with an aryl (*e.g.*, phenylmethyl (benzyl)). The term “alkyl” 15 also includes the side chains of natural and unnatural amino acids. The term “*n*-alkyl” means a straight chain (*i.e.*, unbranched) unsubstituted alkyl group.

The term “alkenyl” includes unsaturated aliphatic groups analogous in length and possible substitution to the alkyls described above, but that contain at least one double bond. For example, the term “alkenyl” includes straight-chain alkenyl groups (*e.g.*, ethylenyl, 20 propenyl, butenyl, pentenyl, hexenyl, heptenyl, octenyl, nonenyl, decenyl, *etc.*), branched-chain alkenyl groups, cycloalkenyl (alicyclic) groups (cyclopropenyl, cyclopentenyl, cyclohexenyl, cycloheptenyl, cyclooctenyl), alkyl or alkenyl substituted cycloalkenyl groups, and cycloalkyl or cycloalkenyl substituted alkenyl groups. In certain embodiments, a straight chain or branched chain alkenyl group has 6 or fewer carbon atoms in its backbone (*e.g.*, C₂- 25 C₆ for straight chain, C₃-C₆ for branched chain). Likewise, cycloalkenyl groups may have from 3-8 carbon atoms in their ring structure, and more preferably have 5 or 6 carbons in the ring structure. The term C₂-C₆ includes alkenyl groups containing 2 to 6 carbon atoms.

Moreover, unless otherwise specified, the term alkenyl includes both “unsubstituted alkenyls” and “substituted alkenyls,” the latter of which refers to alkenyl moieties having substituents replacing a hydrogen on one or more carbons of the hydrocarbon backbone. Such substituents can include, for example, alkyl groups, alkynyl groups, halogens, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, 5 alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, cyano, amino (including alkyl amino, dialkylamino, arylamino, diarylamino, and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, 10 alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

The term “alkynyl” includes unsaturated aliphatic groups analogous in length and possible substitution to the alkyls described above, but which contain at least one triple bond. 15 For example, the term “alkynyl” includes straight-chain alkynyl groups (*e.g.*, ethynyl, propynyl, butynyl, pentynyl, hexynyl, heptynyl, octynyl, nonynyl, decynyl, *etc.*), branched-chain alkynyl groups, and cycloalkyl or cycloalkenyl substituted alkynyl groups. In certain embodiments, a straight chain or branched chain alkynyl group has 6 or fewer carbon atoms in its backbone (*e.g.*, C₂-C₆ for straight chain, C₃-C₆ for branched chain). The term C₂-C₆ 20 includes alkynyl groups containing 2 to 6 carbon atoms.

Moreover, unless otherwise specified, the term alkynyl includes both “unsubstituted alkynyls” and “substituted alkynyls,” the latter of which refers to alkynyl moieties having substituents replacing a hydrogen on one or more carbons of the hydrocarbon backbone. Such substituents can include, for example, alkyl groups, alkynyl groups, halogens, hydroxyl, 25 alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, cyano, amino (including alkyl amino, dialkylamino, arylamino, diarylamino, and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, 30 alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

Unless the number of carbons is otherwise specified, “lower alkyl” as used herein means an alkyl group, as defined above, but having from one to five carbon atoms in its 35 backbone structure. “Lower alkenyl” and “lower alkynyl” have chain lengths of, for example, 2-5 carbon atoms.

The term "alkoxy" includes substituted and unsubstituted alkyl, alkenyl, and alkynyl groups covalently linked to an oxygen atom. Examples of alkoxy groups include methoxy, ethoxy, isopropoxy, propoxy, butoxy, and pentoxy groups. Examples of substituted alkoxy groups include halogenated alkoxy groups. The alkoxy groups can be substituted with groups such as alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxy carbonyloxy, aryloxy carbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxy carbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxy, phosphate, phosphonato, phosphinato, cyano, amino (including alkyl amino, dialkylamino, arylamino, diarylamino, and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moieties. Examples of halogen substituted alkoxy groups include, but are not limited to, fluoromethoxy, difluoromethoxy, trifluoromethoxy, chloromethoxy, dichloromethoxy, trichloromethoxy, *etc.*

The term "heteroatom" includes atoms of any element other than carbon or hydrogen. Preferred heteroatoms are nitrogen, oxygen, sulfur and phosphorus.

The term "hydroxy" or "hydroxyl" includes groups with an -OH or -O⁻ (with an appropriate counterion).

The term "halogen" includes fluorine, bromine, chlorine, iodine, *etc.* The term "perhalogenated" generally refers to a moiety wherein all hydrogens are replaced by halogen atoms.

The term "substituted" includes substituents which can be placed on the moiety and which allow the molecule to perform its intended function. Examples of substituents include alkyl, alkenyl, alkynyl, aryl, (CR'R')₀₋₃NR'R', (CR'R')₀₋₃CN, NO₂, halogen, (CR'R')₀₋₃C(halogen)₃, (CR'R')₀₋₃CH(halogen)₂, (CR'R')₀₋₃CH₂(halogen), (CR'R')₀₋₃CONR'R', (CR'R')₀₋₃S(O)₁₋₂NR'R', (CR'R')₀₋₃CHO, (CR'R')₀₋₃O(CR'R')₀₋₃H, (CR'R')₀₋₃S(O)₀₋₂R', (CR'R')₀₋₃O(CR'R')₀₋₃H, (CR'R')₀₋₃COR', (CR'R')₀₋₃CO₂R', or (CR'R')₀₋₃OR' groups; wherein each R' and R' are each independently hydrogen, a C₁-C₅ alkyl, C₂-C₅ alkenyl, C₂-C₅ alkynyl, or aryl group, or R' and R' taken together are a benzylidene group or a -(CH₂)₂O(CH₂)₂- group.

The term "amine" or "amino" includes compounds or moieties in which a nitrogen atom is covalently bonded to at least one carbon or heteroatom. The term "alkyl amino" includes groups and compounds wherein the nitrogen is bound to at least one additional alkyl group. The term "dialkyl amino" includes groups wherein the nitrogen atom is bound to at least two additional alkyl groups.

The term "ether" includes compounds or moieties which contain an oxygen bonded to two different carbon atoms or heteroatoms. For example, the term includes "alkoxyalkyl" which refers to an alkyl, alkenyl, or alkynyl group covalently bonded to an oxygen atom which is covalently bonded to another alkyl group.

- 5 The term "ester" includes compounds and moieties which contain a carbon or a heteroatom bound to an oxygen atom which is bonded to the carbon of a carbonyl group. The term "ester" includes alkoxycarboxy groups such as methoxycarbonyl, ethoxycarbonyl, propoxycarbonyl, butoxycarbonyl, pentoxycarbonyl, *etc.*

- 10 The term "base" includes the known purine and pyrimidine heterocyclic bases, deazapurines, and analogs (including heterocycl substituted analogs, *e.g.*, aminoethoxy phenoxazine), derivatives (*e.g.*, 1-alkenyl-, 1-alkynyl-, heteroaromatic-, and 1-alkynyl derivatives) and tautomers thereof. Examples of purines include adenine, guanine, inosine, diaminopurine, and xanthine and analogs (*e.g.*, 8-oxo-N⁶-methyladenine or 7-diazaxanthine) and derivatives thereof. Pyrimidines include, for example, thymine, uracil, and cytosine, and
15 their analogs (*e.g.*, 5-methylcytosine, 5-methyluracil, 5-(1-propynyl)uracil, 5-(1-propynyl)cytosine and 4,4-ethanocytosine). Other examples of suitable bases include non-purinyl and non-pyrimidinyl bases such as 2-aminopyridine and triazines.

- The term "nucleoside" includes bases which are covalently attached to a sugar moiety, preferably ribose or deoxyribose. Examples of preferred nucleosides include ribonucleosides
20 and deoxyribonucleosides. Nucleosides also include bases linked to amino acids or amino acid analogs which may comprise free carboxyl groups, free amino groups, or protecting groups. Suitable protecting groups are well known in the art (*see* P.G.M. Wuts and T.W. Greene, "Protective Groups in Organic Synthesis", 2nd Ed., Wiley-Interscience, New York, 1999).

- 25 The term "nucleotide" includes nucleosides which further comprise a phosphate group or a phosphate analog.

 In a preferred embodiment, the nucleomonomers of an oligonucleotide of the invention are RNA nucleotides. In another preferred embodiment, the nucleomonomers of an oligonucleotide of the invention are modified RNA nucleotides.

As used herein, the term “linkage” includes a naturally occurring, unmodified phosphodiester moiety (-O-(PO₂)-O-) that covalently couples adjacent nucleomonomers. As used herein, the term “substitute linkage” includes any analog or derivative of the native phosphodiester group that covalently couples adjacent nucleomonomers. Substitute linkages include phosphodiester analogs, *e.g.*, such as phosphorothioate, phosphorodithioate, and P-ethoxyphosphodiester, P-ethoxyphosphodiester, P-alkyloxyphosphotriester, methylphosphonate, and nonphosphorus containing linkages, *e.g.*, such as acetals and amides. Such substitute linkages are known in the art (*e.g.*, Bjergaard *et al.* 1991. *Nucleic Acids Res.* 19:5843; Caruthers *et al.* 1991. *Nucleosides Nucleotides.* 10:47).

Oligonucleotides of the invention comprise 3' and 5' termini (except for circular oligonucleotides). The 3' and 5' termini of an oligonucleotide can be substantially protected from nucleases *e.g.*, by modifying the 3' or 5' linkages (*e.g.*, U.S. patent 5,849,902 and WO 98/13526). For example, oligonucleotides can be made resistant by the inclusion of a “blocking group.” The term “blocking group” as used herein refers to substituents (*e.g.*, other than OH groups) that can be attached to oligonucleotides or nucleomonomers, either as protecting groups or coupling groups for synthesis (*e.g.*, hydrogen phosphonate, phosphoramidite, or PO₃²⁻). “Blocking groups” also include “end blocking groups” or “exonuclease blocking groups” which protect the 5' and 3' termini of the oligonucleotide, including modified nucleotides and non-nucleotide exonuclease resistant structures.

Exemplary end-blocking groups include cap structures (*e.g.*, a 7-methylguanosine cap), inverted nucleomonomers, *e.g.*, with 3'-3' or 5'-5' end inversions (see *e.g.*, Ortiagao *et al.* 1992. *Antisense Res. Dev.* 2:129), methylphosphonate, phosphoramidite, non-nucleotide groups (*e.g.*, non-nucleotide linkers, amino linkers, conjugates) and the like. The 3' terminal nucleomonomer can comprise a modified sugar moiety. The 3' terminal nucleomonomer can comprise a 3'-O that can optionally be substituted by a blocking group that prevents 3'-exonuclease degradation of the oligonucleotide. For example, the 3'-hydroxyl can be esterified to a nucleotide through a 3'→3' internucleotide linkage. For example, the alkyloxy radical can be methoxy, ethoxy, or isopropoxy, and preferably, ethoxy. Optionally, the 3'→3' linked nucleotide at the 3' terminus can be linked by a substitute linkage. To reduce nuclease degradation, the 5' most 3'→5' linkage can be a modified linkage, *e.g.*, a phosphorothioate or a P-alkyloxyphosphotriester linkage. Preferably, the two 5' most 3'→5' linkages are modified linkages. Optionally, the 5' terminal hydroxy moiety can be esterified with a phosphorus containing moiety, *e.g.*, phosphate, phosphorothioate, or P-ethoxyphosphate.

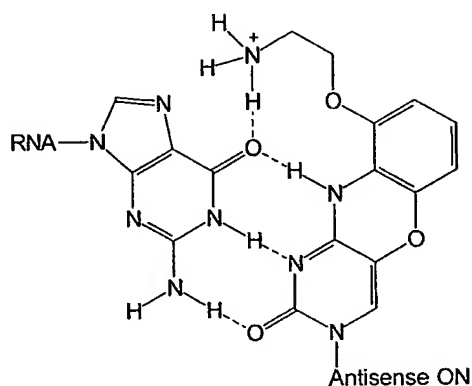
In one embodiment, an oligonucleotide may comprise a 5' phosphate group or a group larger than a phosphate group.

In one embodiment, the oligonucleotides included in the composition are high affinity oligonucleotides. The term "high affinity" as used herein includes oligonucleotides that have a T_m (melting temperature) of or greater than about 60°C, greater than about 65°C, greater than about 70°C, greater than about 75°C, greater than about 80 °C or greater than about 85 °C. The T_m is the midpoint of the temperature range over which the oligonucleotide separates from the target nucleotide sequence. At this temperature, 50% helical (hybridized) versus coil (unhybridized) forms are present. T_m is measured by using the UV spectrum to determine the formation and breakdown (melting) of hybridization. Base stacking occurs during hybridization, which leads to a reduction in UV absorption. T_m depends both on GC content of the two nucleic acid molecules and on the degree of sequence complementarity. T_m can be determined using techniques that are known in the art (see for example, Monia *et al.* 1993. *J. Biol. Chem.* 268:145; Chiang *et al.* 1991. *J. Biol. Chem.* 266:18162; Gagnor *et al.* 1987. *Nucleic Acids Res.* 15:10419; Monia *et al.* 1996. *Proc. Natl. Acad. Sci.* 93:15481; Publis and Tinoco. 1989. *Methods in Enzymology* 180:304; Thuong *et al.* 1987. *Proc. Natl. Acad. Sci. USA* 84:5129).

One skilled in the art will recognize that the length of an RNAi oligonucleotide corresponds to a region of complementarity to the target in the antisense stranded, and the RNAi may be longer, if, for example the RNAi is of a hairpin design.

In one embodiment, an oligonucleotide can include an agent which increases the affinity of the oligonucleotide for its target sequence. The term "affinity enhancing agent" includes agents that increase the affinity of an oligonucleotide for its target. Such agents include, *e.g.*, intercalating agents and high affinity nucleomonomers. Intercalating agents interact strongly and nonspecifically with nucleic acids. Intercalating agents serve to stabilize RNA-DNA duplexes and thus increase the affinity of the oligonucleotides for their targets. Intercalating agents are most commonly linked to the 3' or 5' end of oligonucleotides. Examples of intercalating agents include: acridine, chlorambucil, benzopyridoquinoline, benzopyridoindole, benzophenanthridine, and phenazinium. The agents may also impart other characteristics to the oligonucleotide, for example, increasing resistance to endonucleases and exonucleases.

In one embodiment, a high affinity nucleomonomer is incorporated into an oligonucleotide. The language "high affinity nucleomonomer" as used herein includes modified bases or base analogs that bind to a complementary base in a target nucleic acid molecule with higher affinity than an unmodified base, for example, by having more energetically favorable interactions with the complementary base, *e.g.*, by forming more hydrogen bonds with the complementary base. For example, high affinity nucleomonomer analogs such as aminoethoxy phenoxazine (also referred to as a G clamp), which forms four hydrogen bonds with guanine are included in the term "high affinity nucleomonomer." A high affinity nucleomonomer is illustrated below (*see, e.g., Flanagan, et al., 1999. Proc. Natl. Acad. Sci. 96:3513*).



(*i.e., guanine and aminoethoxy phenoxazine*)

Other exemplary high affinity nucleomonomers are known in the art and include 7-alkenyl, 7-alkynyl, 7-heteroaromatic-, or 7-alkynyl-heteroaromatic-substituted bases or the like which can be substituted for adenosine or guanosine in oligonucleotides (*see e.g., U.S. patent 5,594,121*). Also, 7-substituted deazapurines have been found to impart enhanced binding properties to oligonucleotides, *i.e.*, by allowing them to bind with higher affinity to complementary target nucleic acid molecules as compared to unmodified oligonucleotides. High affinity nucleomonomers can be incorporated into the oligonucleotides of the instant invention using standard techniques.

In another embodiment, an agent that increases the affinity of an oligonucleotide for its target comprises an intercalating agent. As used herein the language "intercalating agent" includes agents which can bind to a DNA double helix. When covalently attached to an oligonucleotide of the invention, an intercalating agent enhances the binding of the
5 oligonucleotide to its complementary genomic DNA target sequence. The intercalating agent may also increase resistance to endonucleases and exonucleases. Exemplary intercalating agents are taught by Helene and Thuong (1989. *Genome* 31:413), and include *e.g.*, acridine derivatives (Lacoste *et al.* 1997. *Nucleic Acids Research*. 25:1991; Kukreti *et al.* 1997. *Nucleic Acids Research*. 25:4264); quinoline derivatives (Wilson *et al.* 1993. *Biochemistry*
10 32:10614); benzo[f]quino[3,4-b]quinoxaline derivatives (Marchand *et al.* 1996. *Biochemistry*. 35:5022; Escude *et al.* 1998. *Proc. Natl. Acad. Sci.* 95:3591). Intercalating agents can be incorporated into an oligonucleotide using any convenient linkage. For example, acridine or psoralen can be linked to the oligonucleotide through any available -OH or -SH group, *e.g.*, at the terminal 5' position of the oligonucleotide, the 2' positions of sugar moieties, or an OH,
15 NH₂, COOH or SH incorporated into the 5-position of pyrimidines using standard methods.

In one embodiment, when included in an RNase H activating antisense oligonucleotide, an agent that increases the affinity of an oligonucleotide for its target is not positioned adjacent to an RNase activating region of the oligonucleotide, *e.g.*, is positioned adjacent to a non-RNase activating region. Preferably, the agent that increases the affinity of
20 an oligonucleotide for its target is placed at a distance as far as possible from the RNase activating domain of the chimeric antisense oligonucleotide such that the specificity of the chimeric antisense oligonucleotide is not altered when compared with the specificity of a chimeric antisense oligonucleotide which lacks the intercalating compound. In one embodiment, this can be accomplished by positioning the agent adjacent to a non-RNase
25 activating region. The specificity of the oligonucleotide can be tested by demonstrating that transcription of a non-target sequence. Preferably a non-target sequence which is structurally similar to the target (*e.g.*, has some sequence homology or identity with the target sequence but which is not identical in sequence to the target) is not inhibited to a greater degree by an oligonucleotide comprising an affinity enhancing agent directed against the target than by an
30 oligonucleotide that does not comprise an affinity enhancing agent that is directed against the target.

In one embodiment, the oligonucleotides of the invention are GC enriched. As used herein the term "GC enriched" includes oligonucleotides that have a relatively high percent GC content. For example, in one embodiment an oligonucleotide of the invention has at least
35 about 20%, at least about 30%, at least about 40% GC content. In another embodiment, an oligonucleotide of the invention has at least about 50%, at least about 60%, or at least about 70% GC content.

In one embodiment, the oligonucleotides of the invention are at least about 25 nucleomonomers in length. In one embodiment, the antisense oligonucleotides of the invention are greater than about 25 nucleomonomers in length. In one embodiment, an antisense oligonucleotide of the invention is at least about 30, at least about 40, at least about 50, or at least about 60, at least about 70, at least about 80, or at least about 90 nucleomonomers in length.

Double-stranded RNA Oligonucleotides

Double-stranded RNA (double-stranded RNA or RNAi (double-stranded RNA interference)) is a double-stranded RNA oligonucleotide that can be used to inhibit protein synthesis in a cell (see, *e.g.*, WO 01/36646A1; Elbashir *et al.* 2001. *Genes & Development* 15:188; Elbashir *et al.* 2001. *Nature* 411:494; Elbashir *et al.* 2001 *EMBO*. 20:6877). Double-stranded RNA may be formed by a single, self-complementary strand or two separate complementary strands. Duplex formation can occur either inside or outside the cell containing the target gene.

As used herein, the term "double-stranded" includes one or more nucleic acid molecules comprising a region of the molecule in which at least a portion of the nucleomonomers are complementary and hydrogen bond to form a duplex.

As used herein, the term "duplex" includes the region of the double-stranded nucleic acid molecule(s) that is (are) hydrogen bonded to a complementary sequence.

Accordingly, one aspect of the invention is a method of inhibiting the activity of a target gene by introducing an RNAi agent into a cell, such that the dsRNA component of the RNAi agent is targeted to the gene. In one embodiment, an RNA oligonucleotide molecule may contain at least one nucleomonomer that is a modified nucleotide analogue. The nucleotide analogues may be located at positions where the target-specific activity, *e.g.*, the RNAi mediating activity is not substantially effected, *e.g.*, in a region at the 5'-end or the 3'-end of the double-stranded molecule, where the overhangs may be stabilized by incorporating modified nucleotide analogues.

In another aspect, double-stranded RNA molecules known in the art can be used in the methods of the present invention. Double-stranded RNA molecules known in the art may also be modified according to the teachings herein in conjunction with such methods, *e.g.*, by using modified nucleomonomers. For example, *see* U.S. 6,506,559; U.S. 2002/0,173,478 A1; 5 U.S. 2002/0,086,356 A1; Shuey, *et al.*, "RNAi: gene-silencing in therapeutic intervention." Drug Discov. Today 2002 Oct 15;7(20):1040-6; Aoki, *et al.*, "Clin. Exp. Pharmacol. Physiol. 2003 Jan;30(1-2):96-102; Cioca, *et al.*, "RNA interference is a functional pathway with therapeutic potential in human myeloid leukemia cell lines. Cancer Gene Ther. 2003 Feb;10(2):125-33.

10 Further examples of double-stranded RNA molecules include those disclosed in the following references: Kawasaki, *et al.*, "Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells." Nucleic Acids Res. 2003 Jan 15;31(2):700-7; Cottrell, *et al.*, "Silence of the strands: RNA interference in eukaryotic pathogens." Trends Microbiol. 2003 Jan;11(1):37- 15 43; Links, "Mammalian RNAi for the masses." Trends Genet. 2003 Jan;19(1):9-12; Hamada, *et al.*, "Effects on RNA interference in gene expression (RNAi) in cultured mammalian cells of mismatches and the introduction of chemical modifications at the 3'-ends of siRNAs." Antisense Nucleic Acid Drug Dev. 2002 Oct;12(5):301-9; Links, "RNAi and related mechanisms and their potential use for therapy." Curr. Opin. Chem. Biol. 2002 Dec;6(6):829- 20 34; Kawasaki, *et al.*, "Short hairpin type of dsRNAs that are controlled by tRNA(Val) promoter significantly induce RNAi-mediated gene silencing in the cytoplasm of human cells." Nucleic Acids Res. 2003 Jan 15;31(2):700-7.)

Double-stranded RNA molecule comprises a nucleotide sequence which is substantially identical to at least part of the target gene. In one embodiment, a double- 25 stranded RNA molecule comprises a nucleotide sequence which is at least about 100 % identical to a portion of the target gene. In another embodiment, a double-stranded RNA molecule comprises a nucleotide sequence which is at least about 95 % identical to a portion of the target gene. In another embodiment, a double-stranded RNA molecule comprises a nucleotide sequence which is at least about 90 % identical to a portion of the target gene. In 30 another embodiment, a double-stranded RNA molecule comprises a nucleotide sequence which is at least about 80 % identical to a portion of the target gene. In another embodiment, a double-stranded RNA molecule comprises a nucleotide sequence which is at least about 60 % identical to a portion of the target gene. In another embodiment, a double-stranded RNA molecule comprises a nucleotide sequence which is at least about 100 % identical to a portion 35 of the target gene.

To determine the percent identity of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the
5 length of the target gene sequence aligned for comparison purposes is at least about 25 nucleotide residues, at least about 50, at least about 100, at least about 150, at least about 200, or at least about 300 or more nucleotide residues are aligned. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then
10 the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two
15 sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two nucleotide sequences is determined using *e.g.*, the GAP program in the GCG software package, using a NWSgapdna. CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two nucleotide sequences is determined using the algorithm of E.
20 Meyers and W. Miller (*Comput. Appl. Biosci.* 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid sequences of the present invention can further be used as a "query sequence" to perform alignments against sequences in public databases. Such searches can
25 be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST
30 programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See, *e.g.*, the NIH internet website.

In one embodiment, the oligonucleotides of the invention are identical to a target nucleic acid sequence over at least about 80% of the length of the oligonucleotide. In another embodiment, oligonucleotides of the invention are identical to a target nucleic acid sequence
35 over at least about 90-95 % of the length of the oligonucleotide. In another embodiment, oligonucleotides of the invention are identical to a target nucleic acid sequence over the entire length of the oligonucleotide.

In yet another embodiment, a sequence of a double-stranded RNA molecule of the invention hybridizes to at least a portion of the target gene under stringent hybridization conditions. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 5 60% complementary to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% complementary to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. 10 (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50°C, preferably at 55°C, more preferably at 60°C, and even more preferably at 65°C. Ranges intermediate to the above-recited values, e.g., at 60-65°C or at 55-60°C are also intended to be encompassed by the present invention. 15 Alternatively, formamide can be included in the hybridization solution, using methods and conditions also known in the art.

Antisense Oligonucleotides

As used herein, the term "antisense oligonucleotide" includes oligonucleotides which 20 comprise a nucleotide sequence which specifically interferes with the synthesis of the target polypeptide. In general, antisense oligonucleotides of the invention bind to the "sense" strand of the nucleotide sequence of the target gene (e.g., polynucleotides such as DNA, mRNA (including pre-mRNA)) molecules. When antisense oligonucleotides of the invention bind to nucleic acid molecules, they can bind to any region of the nucleic acid molecule, 25 including e.g., introns, exons, 5', or 3' untranslated regions. For example, antisense oligonucleotides that work as steric blockers preferentially bind within a splice junction, 5' untranslated region, or the start region of a nucleic acid target molecule. Antisense oligonucleotides that work by activating RNase H preferably bind within an intron, an exon, the 5' untranslated region, or the 3' untranslated region of a nucleic acid target molecule.

Antisense oligonucleotides of the invention may or may not be complementary to their target sequence. Without being limited to any particular mechanism of action, an antisense oligonucleotide used in an oligonucleotide composition of the invention that can specifically hybridize with a nucleotide sequence within the target gene (*i.e.*, is
5 complementary to a nucleotide sequence within the target gene) may achieve its affects based on, *e.g.*, (1) binding to target mRNA and sterically blocking the ribosome complex from translating the mRNA; (2) binding to target mRNA and triggering mRNA cleavage by RNase H; (3) binding to double-stranded DNA in the nucleus and forming a triple helix; (4) hybridizing to open DNA loops created by RNA polymerase; (5) interfering with mRNA
10 splicing; (6) interfering with transport of mRNA from the nucleus to the cytoplasm; or (7) interfering with translation through inhibition of the binding of initiation factors or assembly of ribosomal subunits (*i.e.*, at the start codon).

Without being limited to any particular mechanism of action, the antisense oligonucleotides used in an oligonucleotide composition of the invention that can not
15 specifically hybridize with a nucleotide sequence within the target gene (are not complementary to a nucleotide sequence within the target gene) may achieve their affects based on, *e.g.*, (1) the secondary structure of the oligonucleotide; (2) hybridization to a different nucleotide sequence; (3) binding to proteins or other molecules that may affect the target gene; or (4) modulating oligonucleotide degradation products which themselves can
20 affect cellular functions.

In one embodiment, at least two of the antisense oligonucleotides in an oligonucleotide composition of the invention inhibit protein synthesis via the same mechanism. In another embodiment, at least two of the antisense oligonucleotides in an oligonucleotide composition inhibit protein synthesis via a different mechanism. In yet
25 another embodiment, all of the antisense oligonucleotides present in an oligonucleotide composition inhibit protein synthesis via the same mechanism. The oligonucleotide compositions of the present invention may comprise antisense oligonucleotides which rely simultaneously on several of these modes of action.

The antisense oligonucleotides used in an oligonucleotide composition of the
30 invention may be of any type, *e.g.*, including morpholino oligonucleotides, RNase H activating oligonucleotides, or ribozymes.

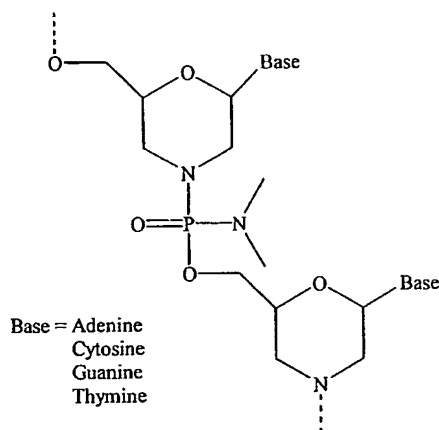
In one embodiment, antisense oligonucleotides of the invention are substantially complementary to a target nucleic acid sequence. Percent complementarity is determined analogously to percent identity. For example, when a position in a test nucleotide sequence is occupied by a nucleotide that is complementary to the corresponding position in the reference sequence, then the molecules are complementary at that position. In one embodiment, an antisense RNA molecule comprises a nucleotide sequence which is at least about 100 % complementary to a portion of the target gene. In another embodiment, an antisense RNA molecule comprises a nucleotide sequence which is at least about 90 % complementary to a portion of the target gene. In another embodiment, an antisense RNA molecule comprises a nucleotide sequence which is at least about 80 % complementary to a portion of the target gene. In another embodiment, an antisense RNA molecule comprises a nucleotide sequence which is at least about 60 % complementary to a portion of the target gene. In another embodiment, an antisense RNA molecule comprises a nucleotide sequence which is at least about 100 % complementary to a portion of the target gene. Preferably, no loops greater than about 8 nucleotides are formed by areas of non-complementarity between the oligonucleotide and the target.

In one embodiment, the antisense oligonucleotides of the invention are complementary to a target nucleic acid sequence over at least about 80% of the length of the oligonucleotide. In another embodiment, antisense oligonucleotides of the invention are complementary to a target nucleic acid sequence over at least about 90-95 % of the length of the oligonucleotide. In another embodiment, antisense oligonucleotides of the invention are complementary to a target nucleic acid sequence over the entire length of the oligonucleotide.

Antisense oligonucleotides of the invention can be "chimeric oligonucleotides" which comprise an RNA-like and a DNA-like region. The language "RNase H activating region" includes a region of an oligonucleotide, *e.g.*, a chimeric oligonucleotide, that is capable of recruiting RNase H to cleave the target RNA strand to which the oligonucleotide binds. Typically, the RNase activating region contains a minimal core (of at least about 3-5, typically between about 3-12, more typically, between about 5-12, and more preferably between about 5-10 contiguous nucleomonomers) of DNA or DNA-like nucleomonomers. (See *e.g.*, US patent 5,849,902). More preferably, the RNase H activating region comprises about nine contiguous deoxyribose containing nucleomonomers. Preferably, the contiguous nucleomonomers are linked by a substitute linkage, *e.g.*, a phosphorothioate linkage.

The language “non-activating region” includes a region of an antisense oligonucleotide, *e.g.*, a chimeric oligonucleotide, that does not recruit or activate RNase H. Preferably, a non-activating region does not comprise phosphorothioate DNA. The oligonucleotides of the invention comprise at least one non-activating region. In one embodiment, the non-activating region can be stabilized against nucleases or can provide specificity for the target by being complementary to the target and forming hydrogen bonds with the target nucleic acid molecule, which is to be bound by the oligonucleotide.

Antisense oligonucleotides of the present invention may include “morpholino oligonucleotides.” Morpholino oligonucleotides are non-ionic and function by an RNase H-independent mechanism. Each of the 4 genetic bases (Adenine, Cytosine, Guanine, and Thymine/Uracil) of the morpholino oligonucleotides is linked to a 6-membered morpholine ring. Morpholino oligonucleotides are made by joining the 4 different subunit types by non-ionic phosphorodiamidate intersubunit linkages. An example of a 2 subunit morpholino oligonucleotide is shown below.



Morpholino oligonucleotides have many advantages including complete resistance to nucleases (Antisense & Nuc. Acid Drug Dev. 1996. 6:267); predictable targeting (Biochemica Biophysica Acta. 1999. 1489:141); reliable activity in cells (Antisense & Nuc. Acid Drug Dev. 1997. 7:63); excellent sequence specificity (Antisense & Nuc. Acid Drug Dev. 1997. 7:151); minimal non-antisense activity (Biochemica Biophysica Acta. 1999. 1489:141); and simple osmotic or scrape delivery (Antisense & Nuc. Acid Drug Dev. 1997. 7:291). Morpholino oligonucleotides are also preferred because of their non-toxicity at high doses. A discussion of the preparation of morpholino oligonucleotides can be found in Antisense & Nuc. Acid Drug Dev. 1997. 7:187.

A variety of nucleotides of different lengths may be used. In one embodiment, an oligonucleotide of the invention is greater than about 25 nucleomonomers in length. In one embodiment, an oligonucleotide of the invention is at least about 10, 12, 14, 16, 18, 20, 22, 24, 26, 27, 28, 29, 30, at least about 40, at least about 50, or at least about 60, at least about
5 70, at least about 80, or at least about 90 nucleomonomers in length. In another embodiment, an oligonucleotide of the invention is less than about 25 nucleomonomers in length, particularly about 21 to 23. In yet another embodiment, an oligonucleotide of the invention is about 10, 12, 14, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleomonomers in length. In another embodiment, an oligonucleotide of the invention is at most about 26, 27,
10 28, 29, 30, at most about 40, at most about 50, or at most about 60, at most about 70, at most about 80, or at most about 90 nucleomonomers in length.

Preferred nucleomonomers in some aspects are ribonucleotides, including 2'-O-methyl ribonucleotides and other 2'-modified RNA molecules.

Oligomers of the invention may also comprise a DNA gap or a phosphorothioate
15 DNA gap.

In some aspects, the present invention relates to compositions and methods comprising at least about 4, 5, 6, 7, 8, 9, or 10 antisense oligonucleotides targeting at least four, five, six, seven, eight, nine, or ten different nucleic acid sequences.

Selection Of Oligonucleotide Sequences

Once the target protein is selected and the nucleotide sequence which encodes it is determined, the sequence of an oligonucleotide for inclusion in the compositions of the invention is determined. The sequence of the target gene is analyzed and oligonucleotides are chosen by a process including both elimination and selection steps. In one embodiment, oligonucleotides which have more than 3 of any nucleotide (A, U, C, or G) occurring consecutively within the oligonucleotide are eliminated. In another embodiment, oligonucleotides having dinucleotide repeats (*e.g.*, AUAU, ACAC, AGAG, UCUC, UGUG, or CGCG) are eliminated. In another embodiment, oligonucleotides are chosen that target nucleotide sequences of the target gene that are preferably at least about 25 nucleotides apart. In another embodiment, oligonucleotides are chosen that comprise between 4 and 10 (inclusive) of each base, such that the base composition of the oligonucleotides is similar. In another embodiment, the percentage of bases in the oligonucleotide which are G or C is greater than 50%. In one embodiment, when oligonucleotides are designed to be complementary to a chosen target sequence, preferably, they are 100% complementary to the target sequence. In another embodiment, an oligonucleotide preferably has greater than 2 mismatches to other, non-target genes. This can be tested by one of ordinary skill in the art, *e.g.*, using available alignment programs and public databases, *e.g.*, the National Institutes of Health internet website.

20

Oligonucleotide Compositions of the Invention

This invention relates to oligonucleotide compositions including more than one individual oligonucleotide molecule. The individual oligonucleotide molecules of the composition target at least one target nucleotide sequence of a single target gene. For example, in one embodiment, at least two of the oligonucleotides present in the composition target the same nucleotide sequence in the same target gene *e.g.*, the oligonucleotides comprise different chemistries but target (*e.g.*, specifically hybridize to) the same sequence of bases in a target nucleic acid molecule. In another embodiment, at least two of the oligonucleotides present in the composition target different nucleotide sequences in the same target gene (*e.g.*, the oligonucleotide composition comprises one oligonucleotide targeting a nucleotide sequence in the promoter of a gene and another oligonucleotide targeting a nucleotide sequence in the portion of the coding sequence of the target nucleic acid molecule or the oligonucleotide composition comprises at least two different oligonucleotides that target two different nucleotide sequences in the coding region of the target nucleic acid molecule).

35

The number of oligonucleotides used in an oligonucleotide composition of the invention can vary from as few as about 2 oligonucleotides to greater than about 20 oligonucleotides. In one embodiment, at least about 3-4 different oligonucleotides are used in the oligonucleotide composition. In another embodiment, at least about 5-6 different
5 oligonucleotides are used in the oligonucleotide composition. In a further embodiment, at least about 7-8 different oligonucleotides are used in the oligonucleotide composition. In one embodiment, greater than about 8 different oligonucleotides are used in an oligonucleotide composition of the invention. In a preferred embodiment, the number of different
10 oligonucleotides in the oligonucleotide composition is chosen so as to use the minimum number of different oligonucleotides that effectively inhibit synthesis of the target protein.

The different oligonucleotides used in an oligonucleotide composition of the invention can each be present at the same concentration or can be present in different concentrations. For example, more desirable oligonucleotides (*e.g.*, those that are more inexpensive or easier to synthesize) may be present at higher concentrations than less
15 desirable oligonucleotides.

Preferably, the oligonucleotides in a composition are either all double-stranded RNA oligonucleotides or all antisense oligonucleotides.

It will be understood that the individual oligonucleotides of the invention can be synthesized to comprise different chemistries. For example, in one embodiment, a
20 composition of the invention can comprise at least one oligonucleotide that is optionally GC enriched. In another embodiment, a composition of the invention comprises at least one oligonucleotide that binds to its target with high affinity. In another exemplary embodiment, a composition of the invention comprises at least one that is at least about 25
nucleomonomers in length. In one embodiment, an oligonucleotide of the invention
25 comprises an oligonucleotide that is GC enriched and binds to its target with high affinity. Thus, as shown by this example, one of skill in the art will recognize that given the teachings of the specification, multiple variations of the individual oligonucleotides present in improved oligonucleotide compositions of the invention can be made.

30 Making Oligonucleotide Compositions

In one embodiment, an individual oligonucleotide is not individually tested for its ability to inhibit protein synthesis prior to its inclusion into a composition of the invention.

In another embodiment, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits protein synthesis by about 20% when tested individually. In another embodiment, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits gene expression by about 30% when tested individually.

5 In another embodiment, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits gene expression by about 40% when tested individually. In another embodiment, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits gene expression by about 50% when tested individually. In another embodiment, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits gene
10 expression by about 60% when tested individually. Preferably, an individual oligonucleotide for inclusion in an oligonucleotide composition inhibits gene expression by less than about 40% when tested individually.

In one embodiment, an oligonucleotide composition of the invention inhibits gene expression to an extent that is greater than the level of inhibition of gene expression achieved
15 by any of the individual oligonucleotides of the oligonucleotide composition acting alone. In another embodiment, the oligonucleotide composition achieves a level of inhibition of protein synthesis the same as or higher than the level of inhibition achieved by the most effective individual oligonucleotide of the composition. In one embodiment, an oligonucleotide composition of the present invention is at least about 80% effective at inhibiting gene
20 expression. In another embodiment, an oligonucleotide composition of the present invention is at least about 90%-95% effective at inhibiting gene expression. In another embodiment, an oligonucleotide composition of the present invention is at least about 99% effective at inhibiting gene expression.

The subject compositions greatly increase the efficiency of the inhibition of protein
25 synthesis because the ability of an individual oligonucleotide to inhibit protein synthesis does not have to be tested prior to its inclusion in an oligonucleotide composition of the invention. Accordingly, only one transfection need be done to effectively inhibit protein synthesis. Thus, in one embodiment, an oligonucleotide composition of the invention is contacted with a cell or population of cells prior to testing the ability of the individual oligonucleotides of the
30 composition to inhibit target gene expression. In another embodiment, an oligonucleotide composition of the invention is contacted with a cell or population of cells subsequent to testing the ability of the individual oligonucleotides of the composition to inhibit target gene expression.

To achieve inhibition of gene expression, an oligonucleotide composition of the invention is contacted with a cell (or cell lysate). In one embodiment, the oligonucleotides of an oligonucleotide composition are contacted with a cell simultaneously. In an alternative embodiment, the oligonucleotides of an oligonucleotide composition can be brought into contact with a cell at different times. For example, at least one of the oligonucleotides can be contacted with a cell at a different time from the other oligonucleotides. In yet another example, each of the oligonucleotides of an oligonucleotide composition is contacted with a cell sequentially so that each of the oligonucleotides of an oligonucleotide composition comes into contact with the cell at a different time. As such, the compositions of the instant invention can be formulated for separate administration of the oligonucleotides. Preferably, a cell is contacted with oligonucleotides of the invention such that the level of inhibition of protein synthesis (*e.g.*, as measured either directly (by measuring the decrease in the amount of the target protein produced) or, for example, by measuring the disappearance of a phenotype associated with the presence of the target protein, by measuring a reduction in the amount of mRNA produced from the target gene, or by measuring an increase in the level of degradation of the mRNA) is greater than that observed when individual nucleotides of the invention are tested individually.

The number of oligonucleotides used to contact a cell can vary from as few as 2 oligonucleotides to greater than about 20 oligonucleotides. In one embodiment, at least about 2-3 different oligonucleotides are contacted with a cell. In another embodiment, at least about 4-5 different oligonucleotides are used to contact the cell. In a further embodiment, at least about 6-7 different oligonucleotides are contacted with a cell.

The ability of an oligonucleotide composition of the invention to inhibit protein synthesis can be measured using techniques which are known in the art, for example, by detecting an inhibition in gene transcription or protein synthesis. For example, Nuclease S1 mapping can be performed. In another example, Northern blot analysis can be used to measure the presence of RNA encoding a particular protein. For example, total RNA can be prepared over a cesium chloride cushion (*see, e.g.*, Aulsebrook *et al.*, eds. 1987. *Current Protocols in Molecular Biology* (Greene & Wiley, New York)). Northern blots can then be made using the RNA and probed (*see, e.g., Id.*) In another example, the level of the specific mRNA produced by the target protein can be measured, *e.g.*, using PCR. In yet another example, Western blots can be used to measure the amount of target protein present. In still another embodiment, a phenotype influenced by the amount of the protein can be detected. Techniques for performing Western blots are well known in the art, *see, e.g.*, Chen *et al. J. Biol. Chem.* 271:28259.

In another example, the promoter sequence of a target gene can be linked to a reporter gene and reporter gene transcription (*e.g.*, as described in more detail below) can be monitored. Alternatively, oligonucleotide compositions that do not target a promoter can be identified by fusing a portion of the target nucleic acid molecule with a reporter gene so that the reporter gene is transcribed. By monitoring a change in the expression of the reporter gene in the presence of the oligonucleotide composition, it is possible to determine the effectiveness of the oligonucleotide composition in inhibiting the expression of the reporter gene. For example, in one embodiment, an effective oligonucleotide composition will reduce the expression of the reporter gene. By incrementally adjusting the concentrations and identities of the oligonucleotides in the oligonucleotide composition and monitoring the resulting change in reporter gene expression, it is possible to optimize the oligonucleotide composition.

A "reporter gene" is a nucleic acid that expresses a detectable gene product, which may be RNA or protein. Detection of mRNA expression may be accomplished by Northern blotting and detection of protein may be accomplished by staining with antibodies specific to the protein. Preferred reporter genes produce a readily detectable product. A reporter gene may be operably linked with a regulatory DNA sequence such that detection of the reporter gene product provides a measure of the transcriptional activity of the regulatory sequence. In preferred embodiments, the gene product of the reporter gene is detected by an intrinsic activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detectable signal based on color, fluorescence, or luminescence. Examples of reporter genes include, but are not limited to, those coding for chloramphenicol acetyl transferase (CAT), luciferase, β -galactosidase and alkaline phosphatase.

One skilled in the art would readily recognize numerous reporter genes suitable for use in the present invention. These include, but are not limited to, chloramphenicol acetyltransferase (CAT), luciferase, human growth hormone (hGH), and beta-galactosidase. Examples of such reporter genes can be found in F. A. Ausubel *et al.*, Eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, (1989). Any gene that encodes a detectable product, *e.g.*, any product having detectable enzymatic activity or against which a specific antibody can be raised, can be used as a reporter gene in the present methods.

One reporter gene system is the firefly luciferase reporter system. (Gould, S. J., and Subramani, S. 1988. Anal. Biochem., 7:404-408 incorporated herein by reference). The luciferase assay is fast and sensitive. In this assay, a lysate of the test cell is prepared and combined with ATP and the substrate luciferin. The encoded enzyme luciferase catalyzes a rapid, ATP dependent oxidation of the substrate to generate a light-emitting product. The total light output is measured and is proportional to the amount of luciferase present over a wide range of enzyme concentrations.

CAT is another frequently used reporter gene system; a major advantage of this system is that it has been extensively validated and is widely accepted as a measure of promoter activity. (Gorman C. M., Moffat, L. F., and Howard, B. H. 1982. Mol. Cell. Biol., 2:1044-1051). In this system, test cells are transfected with CAT expression vectors and incubated with the candidate substance within 2-3 days of the initial transfection. Thereafter, cell extracts are prepared. The extracts are incubated with acetyl CoA and radioactive chloramphenicol. Following the incubation, acetylated chloramphenicol is separated from nonacetylated form by thin layer chromatography. In this assay, the degree of acetylation reflects the CAT gene activity with the particular promoter.

Another suitable reporter gene system is based on immunologic detection of hGH. This system is also quick and easy to use. (Selden, R., Burke-Howie, K. Rowe, M. E., Goodman, H. M., and Moore, D. D. (1986), Mol. Cell, Biol., 6:3173-3179 incorporated herein by reference). The hGH system is advantageous in that the expressed hGH polypeptide is assayed in the media, rather than in a cell extract. Thus, this system does not require the destruction of the test cells. It will be appreciated that the principle of this reporter gene system is not limited to hGH but rather adapted for use with any polypeptide for which an antibody of acceptable specificity is available or can be prepared.

25

Uptake Of Oligonucleotides By Cells

Oligonucleotides and oligonucleotide compositions are contacted with (*i.e.*, brought into contact with, also referred to herein as administered or delivered to) and taken up by one or more cells. The term "cells" includes prokaryotic and eukaryotic cells, preferably vertebrate cells, and, more preferably, mammalian cells. In a preferred embodiment, the oligonucleotide compositions of the invention are contacted with human cells.

30

Oligonucleotide compositions of the invention can be contacted with cells *in vitro* or *in vivo*. Oligonucleotides are taken up by cells at a slow rate by endocytosis, but endocytosed oligonucleotides are generally sequestered and not available, *e.g.*, for hybridization to a target nucleic acid molecule. In one embodiment, cellular uptake can be facilitated by
5 electroporation or calcium phosphate precipitation. However, these procedures are only useful for *in vitro* or *ex vivo* embodiments, are not convenient and, in some cases, are associated with cell toxicity.

In another embodiment, delivery of oligonucleotides into cells can be enhanced by suitable art recognized methods including calcium phosphate, DMSO, glycerol or dextran,
10 electroporation, or by transfection, *e.g.*, using cationic, anionic, or neutral lipid compositions or liposomes using methods known in the art (see *e.g.*, WO 90/14074; WO 91/16024; WO 91/17424; U.S. Patent No. 4,897,355; Bergan *et al.* 1993. *Nucleic Acids Research*. 21:3567). Enhanced delivery of oligonucleotides can also be mediated by the use of viruses, polyamine or polycation conjugates using compounds such as polylysine, protamine, or N1, N12-bis
15 (ethyl) spermine (see *e.g.*, Bartzatt, R. *et al.* 1989. *Biotechnol. Appl. Biochem.* 11:133; Wagner E. *et al.* 1992. *Proc. Natl. Acad. Sci.* 88:4255)

Conjugating Agents

Conjugating agents bind to the oligonucleotide in a covalent manner. In one
20 embodiment, oligonucleotides can be derivitized or chemically modified by binding to a conjugating agent to facilitate cellular uptake. For example, covalent linkage of a cholesterol moiety to an oligonucleotide can improve cellular uptake by 5- to 10- fold which in turn improves DNA binding by about 10- fold (Boutorin *et al.*, 1989, *FEBS Letters* 254:129-132). Conjugation of octyl, dodecyl, and octadecyl residues enhances cellular uptake by 3-, 4-, and
25 10- fold as compared to unmodified oligonucleotides (Vlassov *et al.*, 1994, *Biochimica et Biophysica Acta* 1197:95-108). Similarly, derivatization of oligonucleotides with poly-L-lysine can aid oligonucleotide uptake by cells (Schell, 1974, *Biochem. Biophys. Acta* 340:323, and Lemaitre *et al.*, 1987, *Proc. Natl. Acad. Sci. USA* 84:648).

Certain protein carriers can also facilitate cellular uptake of oligonucleotides, including, for example, serum albumin, nuclear proteins possessing signals for transport to the nucleus, and viral or bacterial proteins capable of cell membrane penetration. Therefore, protein carriers are useful when associated with or linked to the oligonucleotides.

- 5 Accordingly, the present invention provides for derivatization of oligonucleotides with groups capable of facilitating cellular uptake, including hydrocarbons and non-polar groups, cholesterol, long chain alcohols (*i.e.*, hexanol), poly-L-lysine and proteins, as well as other aryl or steroid groups and polycations having analogous beneficial effects, such as phenyl or naphthyl groups, quinoline, anthracene or phenanthracene groups, fatty acids, fatty alcohols
10 and sesquiterpenes, diterpenes and steroids. A major advantage of using conjugating agents is to increase the initial membrane interaction that leads to a greater cellular accumulation of oligonucleotides.

Encapsulating Agents

- 15 Encapsulating agents entrap oligonucleotides within vesicles. In another embodiment, an oligonucleotide may be associated with a carrier or vehicle, *e.g.*, liposomes or micelles, although other carriers could be used, as would be appreciated by one skilled in the art. Liposomes are vesicles made of a lipid bilayer having a structure similar to biological membranes. Such carriers are used to facilitate the cellular uptake or targeting of the
20 oligonucleotide, or improve the oligonucleotide's pharmacokinetic or toxicologic properties.

- For example, the oligonucleotides of the present invention may also be administered encapsulated in liposomes, pharmaceutical compositions wherein the active ingredient is contained either dispersed or variously present in corpuscles consisting of aqueous concentric layers adherent to lipidic layers. The oligonucleotides, depending upon solubility, may be
25 present both in the aqueous layer and in the lipidic layer, or in what is generally termed a liposomic suspension. The hydrophobic layer, generally but not exclusively, comprises phospholipids such as lecithin and sphingomyelin, steroids such as cholesterol, more or less ionic surfactants such as diacetylphosphate, stearylamine, or phosphatidic acid, or other materials of a hydrophobic nature. The diameters of the liposomes generally range from about
30 15 nm to about 5 microns.

The use of liposomes as drug delivery vehicles offers several advantages. Liposomes increase intracellular stability, increase uptake efficiency and improve biological activity. Liposomes are hollow spherical vesicles composed of lipids arranged in a similar fashion as those lipids which make up the cell membrane. They have an internal aqueous space for entrapping water soluble compounds and range in size from 0.05 to several microns in diameter. Several studies have shown that liposomes can deliver nucleic acids to cells and that the nucleic acids remain biologically active. For example, a liposome delivery vehicle originally designed as a research tool, such as Lipofectin, can deliver intact nucleic acid molecules to cells.

Specific advantages of using liposomes include the following: they are non-toxic and biodegradable in composition; they display long circulation half-lives; and recognition molecules can be readily attached to their surface for targeting to tissues. Finally, cost-effective manufacture of liposome-based pharmaceuticals, either in a liquid suspension or lyophilized product, has demonstrated the viability of this technology as an acceptable drug delivery system.

Complexing Agents

Complexing agents bind to the oligonucleotide by a strong but non-covalent attraction (*e.g.*, an electrostatic, van der Waals, pi-stacking interaction, etc.). In one embodiment, oligonucleotides of the invention can be complexed with a complexing agent to increase cellular uptake of oligonucleotides. An example of a complexing agent includes cationic lipids. Cationic lipids can be used to deliver oligonucleotides to cells.

The term "cationic lipid" includes lipids and synthetic lipids having both polar and non-polar domains and which are capable of being positively charged at or around physiological pH and which bind to polyanions, such as nucleic acids, and facilitate the delivery of nucleic acids into cells. In general cationic lipids include saturated and unsaturated alkyl and alicyclic ethers and esters of amines, amides, or derivatives thereof. Straight-chain and branched alkyl and alkenyl groups of cationic lipids can contain, *e.g.*, from 1 to about 25 carbon atoms. Preferred straight chain or branched alkyl or alkene groups have six or more carbon atoms. Alicyclic groups include cholesterol and other steroid groups. Cationic lipids can be prepared with a variety of counterions (anions) including, *e.g.*, Cl⁻, Br⁻, I⁻, F⁻, acetate, trifluoroacetate, sulfate, nitrite, and nitrate.

Examples of cationic lipids include: polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, Lipofectin (a combination of DOTMA and DOPE), Lipofectase, Lipofectamine, DOPE, Cytofectin (Gilead Sciences, Foster City, CA), and Eufectins (JBL, San Luis Obispo, CA). Cationic liposomes may comprise the following: N-[1-(2,3-dioleloxy)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1-(2,3-dioleloxy)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3 β -[N-(N',N'-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate (DOSPA), 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB). The cationic lipid N-(1-(2,3-dioleyloxy)propyl)-N,N,N-trimethylammonium chloride (DOTMA), for example, was found to increase 1000-fold the antisense effect of a phosphorothioate oligonucleotide. (Vlassov *et al.*, 1994, *Biochimica et Biophysica Acta* 1197:95-108). Oligonucleotides can also be complexed with, *e.g.*, poly (L-lysine) or avidin and lipids may, or may not, be included in this mixture (*e.g.*, steryl-poly (L-lysine)).

Cationic lipids have been used in the art to deliver oligonucleotides to cells (*see, e.g.*, U.S. Patents 5,855,910; 5,851,548; 5,830,430; 5,780,053; 5,767,099; Lewis *et al.* 1996. *Proc. Natl. Acad. Sci. USA* 93:3176; Hope *et al.* 1998. *Molecular Membrane Biology* 15:1). Other lipid compositions which can be used to facilitate uptake of the instant oligonucleotides can be used in connection with the claimed methods. In addition to those listed supra, other lipid compositions are also known in the art and include, *e.g.*, those taught in U.S. patent 4,235,871; U.S. patents 4,501,728; 4,837,028; 4,737,323.

In one embodiment lipid compositions can further comprise agents, *e.g.*, viral proteins to enhance lipid-mediated transfections of oligonucleotides (Kamata *et al.* 1994. *Nucl. Acids. Res.* 22:536). In another embodiment, oligonucleotides are contacted with cells as part of a composition comprising an oligonucleotide, a peptide, and a lipid as taught, *e.g.*, in U.S. patent 5,736,392. Improved lipids have also been described which are serum resistant (Lewis *et al.* 1996. *Proc. Natl. Acad. Sci.* 93:3176). Cationic lipids and other complexing agents act to increase the number of oligonucleotides carried into the cell through endocytosis.

In another embodiment N-substituted glycine oligonucleotides (peptoids) can be used to optimize uptake of oligonucleotides. Peptoids have been used to create cationic lipid-like compounds for transfection (Murphy *et al.* 1998. *Proc. Natl. Acad. Sci.* 95:1517). Peptoids can be synthesized using standard methods (*e.g.*, Zuckermann, R. N., *et al.* 1992. *J. Am. Chem. Soc.* 114:10646; Zuckermann, R.N., *et al.* 1992. *Int. J. Peptide Protein Res.* 40:497). Combinations of cationic lipids and peptoids, liptoids, can also be used to optimize uptake of the subject oligonucleotides (Hunag *et al.* 1998. *Chemistry and Biology.* 5:345). Liptoids can be synthesized by elaborating peptoid oligonucleotides and coupling the amino terminal submonomer to a lipid via its amino group (Hunag *et al.* 1998. *Chemistry and Biology.* 5:345).

It is known in the art that positively charged amino acids can be used for creating highly active cation lipids (Lewis *et al.* 1996. *Proc. Natl. Acad. Sci. U.S.A.* 93:3176). In one embodiment, a composition for delivering oligonucleotides of the invention comprises a number of arginine, lysine, histidine or ornithine residues linked to a lipophilic moiety (*see, e.g.*, U.S. patent 5,777,153).

In another, a composition for delivering oligonucleotides of the invention comprises a peptide having from between about one to about four basic residues. These basic residues can be located, *e.g.*, on the amino terminal, C-terminal, or internal region of the peptide. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine (can also be considered non-polar), asparagine, glutamine, serine, threonine; tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Apart from the basic amino acids, a majority or all of the other residues of the peptide can be selected from the non-basic amino acids, *e.g.*, amino acids other than lysine, arginine, or histidine. Preferably a preponderance of neutral amino acids with long neutral side chains are used. For example, a peptide such as (N-term) His-Ile-Trp-Leu-Ile-Tyr-Leu-Trp-Ile-Val- (C-term) (SEQ ID NO: ##) could be used. In one embodiment such a composition can be mixed with the fusogenic lipid DOPE as is well known in the art.

In one embodiment, the cells to be contacted with an oligonucleotide composition are contacted with a mixture comprising the oligonucleotide and a mixture comprising a lipid, *e.g.*, one of the lipids or lipid compositions described supra for between about 1 and about five days. In one embodiment, the cells are contacted with a mixture comprising a lipid and the oligonucleotide for between about three days to as long as about 30 days. In another embodiment, a mixture comprising a lipid is left in contact with the cells for at least about five to about 20 days. In another embodiment, a mixture comprising a lipid is left in contact with the cells for at least about seven to about 15 days.

For example, in one embodiment, an oligonucleotide composition can be contacted with cells in the presence of a lipid such as cytofectin CS or GSV(available from Glen Research; Sterling, VA), GS3815, GS2888 for prolonged incubation periods as described herein.

In one embodiment the incubation of the cells with the mixture comprising a lipid and an oligonucleotide composition does not reduce the viability of the cells. Preferably, after the transfection period the cells are substantially viable. In one embodiment, after transfection, the cells are between at least about 70 and at least about 100 percent viable. In another embodiment, the cells are between at least about 80 and at least about 95% viable. In yet another embodiment, the cells are between at least about 85% and at least about 90% viable.

In one embodiment, oligonucleotides are modified by attaching a peptide sequence that transports the oligonucleotide into a cell, referred to herein as a "transporting peptide." In one embodiment, the composition includes an oligonucleotide which is complementary to a target nucleic acid molecule encoding the protein, and a covalently attached transporting peptide.

The language "transporting peptide" includes an amino acid sequence that facilitates the transport of an oligonucleotide into a cell. Exemplary peptides which facilitate the transport of the moieties to which they are linked into cells are known in the art, and include, *e.g.*, HIV TAT transcription factor, lactoferrin, Herpes VP22 protein, and fibroblast growth factor 2 (Pooga *et al.* 1998. *Nature Biotechnology*. 16:857; and Derossi *et al.* 1998. *Trends in Cell Biology*. 8:84; Elliott and O'Hare. 1997. *Cell* 88:223).

For example, in one embodiment, the transporting peptide comprises an amino acid sequence derived from the antennapedia protein. Preferably, the peptide comprises amino acids 43-58 of the antennapedia protein (Arg-Gln-Ile-Lys-Ile-Trp-Phe-Gln-Asn-Arg-Arg-Met-Lys-Trp-Lys-Lys) (SEQ ID NO: ##) or a portion or variant thereof that facilitates transport of an oligonucleotide into a cell (see, *e.g.*, WO 91/1898; Derossi *et al.* 1998. *Trends Cell Biol.* 8:84). Exemplary variants are shown in Derossi *et al.*, *supra*.

In one embodiment, the transporting peptide comprises an amino acid sequence derived from the transportan, galanin (1-12)-Lys-mastoparan (1-14) amide, protein. (Pooga *et al.* 1998. *Nature Biotechnology* 16:857). Preferably, the peptide comprises the amino acids of the transportan protein shown in the sequence GWTLNSAGYLLGKINLKA-
 5 LAALAKKIL (SEQ ID NO: ##) or a portion or variant thereof that facilitates transport of an oligonucleotide into a cell.

In one embodiment, the transporting peptide comprises an amino acid sequence derived from the HIV TAT protein. Preferably, the peptide comprises amino acids 37-72 of the HIV TAT protein, *e.g.*, shown in the sequence C(Acm)FITKALGISYGRKKRRQR-
 10 RRPPQC (SEQ ID NO: ##) (TAT 37-60; where C(Acm) is Cys-acetamidomethyl) or a portion or variant thereof, *e.g.*, C(Acm)GRKKRRQRRPPQC (SEQ ID NO: ##) (TAT 48-40) or C(Acm)LGISYGRKKRRQRRPPQC (SEQ ID NO: ##) (TAT 43-60) that facilitates transport of an oligonucleotide into a cell (Vives *et al.* 1997. *J. Biol. Chem.* 272:16010). In another embodiment the peptide (G)CFITKALGISYGRKKRR-
 15 QRRRPPQGSQTHQVLSLKQ (SEQ ID NO: ##) can be used.

Portions or variants of transporting peptides can be readily tested to determine whether they are equivalent to these peptide portions by comparing their activity to the activity of the native peptide, *e.g.*, their ability to transport fluorescently labeled oligonucleotides to cells. Fragments or variants that retain the ability of the native
 20 transporting peptide to transport an oligonucleotide into a cell are functionally equivalent and can be substituted for the native peptides.

Oligonucleotides can be attached to the transporting peptide using known techniques (*e.g.*, Prochiantz, A. 1996. *Curr. Opin. Neurobiol.* 6:629; Derossi *et al.* 1998. *Trends Cell Biol.* 8:84; Troy *et al.* 1996. *J. Neurosci.* 16:253), Vives *et al.* 1997. *J. Biol. Chem.* 272:16010). For example, in one embodiment, oligonucleotides bearing an activated thiol
 25 group are linked via that thiol group to a cysteine present in a transport peptide (*e.g.*, to the cysteine present in the β turn between the second and the third helix of the antennapedia homeodomain as taught, *e.g.*, in Derossi *et al.* 1998. *Trends Cell Biol.* 8:84; Prochiantz. 1996. *Current Opinion in Neurobiol.* 6:629; Allinquant *et al.* 1995. *J. Cell Biol.* 128:919). In
 30 another embodiment, a Boc-Cys-(Npys)OH group can be coupled to the transport peptide as the last (N-terminal) amino acid and an oligonucleotide bearing an SH group can be coupled to the peptide (Troy *et al.* 1996. *J. Neurosci.* 16:253).

In one embodiment, a linking group can be attached to a nucleomonomer and the transporting peptide can be covalently attached to the linker. In one embodiment, a linker can function as both an attachment site for a transporting peptide and can provide stability against nucleases. Examples of suitable linkers include substituted or unsubstituted C₁-C₂₀ alkyl chains, C₁-C₂₀ alkenyl chains, C₁-C₂₀ alkynyl chains, peptides, and heteroatoms (*e.g.*, S, O, NH, etc.). Other exemplary linkers include bifunctional crosslinking agents such as sulfosuccinimidyl-4-(maleimidophenyl)-butyrate (SMPB) (*see, e.g.*, Smith *et al.* Biochem J 1991. 276: 417-2).

In one embodiment, oligonucleotides of the invention are synthesized as molecular conjugates which utilize receptor-mediated endocytotic mechanisms for delivering genes into cells (*see, e.g.*, Bunnell *et al.* 1992. *Somatic Cell and Molecular Genetics*. 18:559 and the references cited therein).

Targeting Agents

The delivery of oligonucleotides can also be improved by targeting the oligonucleotides to a cellular receptor. The targeting moieties can be conjugated to the oligonucleotides or attached to a carrier group (*i.e.*, poly(L-lysine) or liposomes) linked to the oligonucleotides. This method is well suited to cells that display specific receptor-mediated endocytosis.

For instance, oligonucleotide conjugates to 6-phosphomannosylated proteins are internalized 20-fold more efficiently by cells expressing mannose 6-phosphate specific receptors than free oligonucleotides. The oligonucleotides may also be coupled to a ligand for a cellular receptor using a biodegradable linker. In another example, the delivery construct is mannosylated streptavidin which forms a tight complex with biotinylated oligonucleotides. Mannosylated streptavidin was found to increase 20-fold the internalization of biotinylated oligonucleotides. (Vlassov *et al.* 1994. *Biochimica et Biophysica Acta* 1197:95-108).

In addition specific ligands can be conjugated to the polylysine component of polylysine-based delivery systems. For example, transferrin-polylysine, adenovirus-polylysine, and influenza virus hemagglutinin HA-2 N-terminal fusogenic peptides-polylysine conjugates greatly enhance receptor-mediated DNA delivery in eucaryotic cells. Mannosylated glycoprotein conjugated to poly(L-lysine) in alveolar macrophages has been employed to enhance the cellular uptake of oligonucleotides. Liang *et al.* 1999. *Pharmazie* 54:559-566.

Because malignant cells have an increased need for essential nutrients such as folic acid and transferrin, these nutrients can be used to target oligonucleotides to cancerous cells. For example, when folic acid is linked to poly(L-lysine) enhanced oligonucleotide uptake is seen in promyelocytic leukaemia (HL-60) cells and human melanoma (M-14) cells. Ginobbi
5 *et al.* 1997. *Anticancer Res.* 17:29. In another example, liposomes coated with maleylated bovine serum albumin, folic acid, or ferric protoporphyrin IX, show enhanced cellular uptake of oligonucleotides in murine macrophages, KB cells, and 2.2.15 human hepatoma cells. Liang *et al.* 1999. *Pharmazie* 54:559-566.

Liposomes are naturally targeted to the liver, spleen, and reticuloendothelial system.
10 By coupling liposomes to various ligands such as antibodies or protein A, they can be targeted to specific cell populations. For example, protein A-bearing liposomes may be pretreated with H-2K specific antibodies which are targeted to the mouse major histocompatibility complex-encoded H-2K protein expressed on L cells. (Vlassov *et al.* 1994. *Biochimica et Biophysica Acta* 1197:95-108).

15

Assays of Oligonucleotide Stability

Preferably, the oligonucleotides of the invention are stabilized, *i.e.*, substantially resistant to endonuclease and exonuclease degradation. An oligonucleotide is defined as being substantially resistant to nucleases when it is at least about 3-fold more resistant to
20 attack by an endogenous cellular nuclease, and is highly nuclease resistant when it is at least about 6-fold more resistant than a corresponding, unmodified oligonucleotide. This can be demonstrated by showing that the oligonucleotides of the invention are substantially resist nucleases using techniques which are known in the art.

One way in which substantial stability can be demonstrated is showing that the
25 oligonucleotides of the invention function when delivered to a cell, *e.g.*, that they reduce transcription or translation of target nucleic acid molecules, *e.g.*, by measuring protein levels or by measuring cleavage of mRNA. Assays which measure the stability of target RNA can be performed at about 24 hours post-transfection (*e.g.*, using Northern blot techniques, RNase Protection Assays, or QC-PCR assays as known in the art). Alternatively, levels of the target
30 protein can be measured. Preferably, in addition to testing the RNA or protein levels of interest, the RNA or protein levels of a control, non-targeted gene will be measured (*e.g.*, actin, or preferably a control with sequence similarity to the target) as a specificity control. RNA or protein measurements can be made using any art-recognized technique. Preferably, measurements will be made beginning at about 16-24 hours post transfection. (M. Y. Chiang,
35 *et al.* 1991. *J Biol Chem.* 266:18162-71; T. Fisher, *et al.* 1993. *Nucleic Acids Research.* 21 3857).

The ability of an oligonucleotide composition of the invention to inhibit protein synthesis can be measured using techniques which are known in the art, for example, by detecting an inhibition in gene transcription or protein synthesis. For example, Nuclease S1 mapping can be performed. In another example, Northern blot analysis can be used to
5 measure the presence of RNA encoding a particular protein. For example, total RNA can be prepared over a cesium chloride cushion (*see, e.g.,* Ausubel *et al.*, 1987. Current Protocols in Molecular Biology (Greene & Wiley, New York)). Northern blots can then be made using the RNA and probed (*see, e.g.,* Id.). In another example, the level of the specific mRNA produced by the target protein can be measured, *e.g.,* using PCR. In yet another example,
10 Western blots can be used to measure the amount of target protein present. In still another embodiment, a phenotype influenced by the amount of the protein can be detected. Techniques for performing Western blots are well known in the art, *see, e.g.,* Chen *et al. J. Biol. Chem.* 271:28259.

In another example, the promoter sequence of a target gene can be linked to a reporter
15 gene and reporter gene transcription (*e.g.,* as described in more detail below) can be monitored. Alternatively, oligonucleotide compositions that do not target a promoter can be identified by fusing a portion of the target nucleic acid molecule with a reporter gene so that the reporter gene is transcribed. By monitoring a change in the expression of the reporter gene in the presence of the oligonucleotide composition, it is possible to determine the
20 effectiveness of the oligonucleotide composition in inhibiting the expression of the reporter gene. For example, in one embodiment, an effective oligonucleotide composition will reduce the expression of the reporter gene.

A "reporter gene" is a nucleic acid that expresses a detectable gene product, which may be RNA or protein. Detection of mRNA expression may be accomplished by Northern
25 blotting and detection of protein may be accomplished by staining with antibodies specific to the protein. Preferred reporter genes produce a readily detectable product. A reporter gene may be operably linked with a regulatory DNA sequence such that detection of the reporter gene product provides a measure of the transcriptional activity of the regulatory sequence. In preferred embodiments, the gene product of the reporter gene is detected by an intrinsic
30 activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detectable signal based on color, fluorescence, or luminescence. Examples of reporter genes include, but are not limited to, those coding for chloramphenicol acetyl transferase (CAT), luciferase, β -galactosidase, and alkaline phosphatase.

One skilled in the art would readily recognize numerous reporter genes suitable for use in the present invention. These include, but are not limited to, chloramphenicol acetyltransferase (CAT), luciferase, human growth hormone (hGH), and beta-galactosidase. Examples of such reporter genes can be found in F. A. Ausubel *et al.*, Eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, (1989). Any gene that encodes a detectable product, *e.g.*, any product having detectable enzymatic activity or against which a specific antibody can be raised, can be used as a reporter gene in the present methods.

One reporter gene system is the firefly luciferase reporter system. (Gould, S. J., and Subramani, S. 1988. *Anal. Biochem.*, 7:404-408 incorporated herein by reference). The luciferase assay is fast and sensitive. In this assay, a lysate of the test cell is prepared and combined with ATP and the substrate luciferin. The encoded enzyme luciferase catalyzes a rapid, ATP dependent oxidation of the substrate to generate a light-emitting product. The total light output is measured and is proportional to the amount of luciferase present over a wide range of enzyme concentrations.

CAT is another frequently used reporter gene system; a major advantage of this system is that it has been extensively validated and is widely accepted as a measure of promoter activity. (Gorman C. M., Moffat, L. F., and Howard, B. H. 1982. *Mol. Cell. Biol.*, 2:1044-1051). In this system, test cells are transfected with CAT expression vectors and incubated with the candidate substance within 2-3 days of the initial transfection. Thereafter, cell extracts are prepared. The extracts are incubated with acetyl CoA and radioactive chloramphenicol. Following the incubation, acetylated chloramphenicol is separated from nonacetylated form by thin layer chromatography. In this assay, the degree of acetylation reflects the CAT gene activity with the particular promoter.

Another suitable reporter gene system is based on immunologic detection of hGH. This system is also quick and easy to use. (Selden, R., Burke-Howie, K. Rowe, M. E., Goodman, H. M., and Moore, D. D. (1986), *Mol. Cell, Biol.*, 6:3173-3179 incorporated herein by reference). The hGH system is advantageous in that the expressed hGH polypeptide is assayed in the media, rather than in a cell extract. Thus, this system does not require the destruction of the test cells. It will be appreciated that the principle of this reporter gene system is not limited to hGH but rather adapted for use with any polypeptide for which an antibody of acceptable specificity is available or can be prepared.

Oligonucleotide Synthesis

Oligonucleotides of the invention can be synthesized by any methods known in the art, *e.g.*, using enzymatic synthesis and chemical synthesis. The oligonucleotides can be synthesized *in vitro* (*e.g.*, using enzymatic synthesis and chemical synthesis) or *in vivo* (using recombinant DNA technology well known in the art.

In a preferred embodiment, chemical synthesis is used. Chemical synthesis of linear oligonucleotides is well known in the art and can be achieved by solution or solid phase techniques. Preferably, synthesis is by solid phase methods. Oligonucleotides can be made by any of several different synthetic procedures including the phosphoramidite, phosphite triester, H-phosphonate, and phosphotriester methods, typically by automated synthesis methods.

Oligonucleotide synthesis protocols are well known in the art and can be found, *e.g.*, in U.S. patent 5,830,653; WO 98/13526; Stec *et al.* 1984. *J. Am. Chem. Soc.* 106:6077; Stec *et al.* 1985. *J. Org. Chem.* 50:3908; Stec *et al.* *J. Chromatog.* 1985. 326:263; LaPlanche *et al.* 1986. *Nuc. Acid. Res.* 1986. 14:9081; Fasman G. D., 1989. *Practical Handbook of Biochemistry and Molecular Biology*. 1989. CRC Press, Boca Raton, Fla.; Lamone. 1993. *Biochem. Soc. Trans.* 21:1; U.S. Patent 5,013,830; U.S. Patent 5,214,135; U.S. Patent 5,525,719; Kawasaki *et al.* 1993. *J. Med. Chem.* 36:831; WO 92/03568; U.S. Patent 5,276,019; U.S. Patent 5,264,423.

The synthesis method selected can depend on the length of the desired oligonucleotide and such choice is within the skill of the ordinary artisan. For example, the phosphoramidite and phosphite triester method produce oligonucleotides having 175 or more nucleotides while the H-phosphonate method works well for oligonucleotides of less than 100 nucleotides. If modified bases are incorporated into the oligonucleotide, and particularly if modified phosphodiester linkages are used, then the synthetic procedures are altered as needed according to known procedures. In this regard, Uhlmann *et al.* (1990, *Chemical Reviews* 90:543-584) provide references and outline procedures for making oligonucleotides with modified bases and modified phosphodiester linkages. Other exemplary methods for making oligonucleotides are taught in Sonveaux. 1994. "Protecting Groups in Oligonucleotide Synthesis"; Agrawal. *Methods in Molecular Biology* 26:1. Exemplary synthesis methods are also taught in "Oligonucleotide Synthesis- A Practical Approach" (Gait, M.J. IRL Press at Oxford University Press. 1984). Moreover, linear oligonucleotides of defined sequence can be purchased commercially.

The oligonucleotides may be purified by polyacrylamide gel electrophoresis, or by any of a number of chromatographic methods, including gel chromatography and high pressure liquid chromatography. To confirm a nucleotide sequence, oligonucleotides may be subjected to DNA sequencing by any of the known procedures, including Maxam and Gilbert sequencing, Sanger sequencing, capillary electrophoresis sequencing the wandering spot sequencing procedure or by using selective chemical degradation of oligonucleotides bound to Hybond paper. Sequences of short oligonucleotides can also be analyzed by laser desorption mass spectroscopy or by fast atom bombardment (McNeal, *et al.*, 1982, *J. Am. Chem. Soc.* 104:976; Viari, *et al.*, 1987, *Biomed. Environ. Mass Spectrom.* 14:83; Grotjahn *et al.*, 1982, *Nuc. Acid Res.* 10:4671). Sequencing methods are also available for RNA oligonucleotides.

The quality of oligonucleotides synthesized can be verified by testing the oligonucleotide by capillary electrophoresis and denaturing strong anion HPLC (SAX-HPLC) using, *e.g.*, the method of Bergot and Egan. 1992. *J. Chrom.* 599:35.

Other exemplary synthesis techniques are well known in the art (see, *e.g.*, Sambrook *et al.*, *Molecular Cloning: a Laboratory Manual*, Second Edition (1989); *DNA Cloning*, Volumes I and II (DN Glover Ed. 1985); *Oligonucleotide Synthesis* (MJ Gait Ed, 1984; *Nucleic Acid Hybridisation* (BD Hames and SJ Higgins eds. 1984); *A Practical Guide to Molecular Cloning* (1984); or the series, *Methods in Enzymology* (Academic Press, Inc.)).

Uses of Oligonucleotides

This invention also features methods of inhibiting expression of a protein in a cell including contacting the cell with one of the above-described oligonucleotide compositions.

The oligonucleotides of the invention can be used in a variety of *in vitro* and *in vivo* situations to specifically inhibit protein expression. The instant methods and compositions are suitable for both *in vitro* and *in vivo* use.

In one embodiment, the oligonucleotides of the invention can be used to inhibit gene function *in vitro* in a method for identifying the functions of genes. In this manner, the transcription of genes that are identified, but for which no function has yet been shown, can be inhibited to thereby determine how the phenotype of a cell is changed when the gene is not transcribed. Such methods are useful for the validation of genes as targets for clinical treatment, *e.g.*, with oligonucleotides or with other therapies.

To determine the effect of a composition of the invention, a variety of end points can be used. In addition to the assays described previously herein, for example, nucleic acid probes (*e.g.*, in the form of arrays) can be used to evaluate transcription patterns produced by cells. Probes can also be used detect peptides, proteins, or protein domains, *e.g.*, antibodies
5 can be used to detect the expression of a particular protein. In yet another embodiment, the function of a protein (*e.g.*, enzymatic activity) can be measured. In yet another embodiment, the phenotype of a cell can be evaluated to determine whether or not a target protein is expressed. For example, the ability of a composition to affect a phenotype of a cell that is associated with cancer can be tested.

10 In one embodiment, one or more additional agents (*e.g.*, activating agents, inducing agents, proliferation enhancing agents, tumor promoters) can be added to the cells.

In another embodiment, the compositions of the invention can be used to monitor biochemical reactions such as, *e.g.*, interactions of proteins, nucleic acids, small molecules, or the like--for example the efficiency or specificity of interactions between antigens and
15 antibodies; or of receptors (such as purified receptors or receptors bound to cell membranes) and their ligands, agonists or antagonists; or of enzymes (such as proteases or kinases) and their substrates, or increases or decreases in the amount of substrate converted to a product; as well as many others. Such biochemical assays can be used to characterize properties of the probe or target, or as the basis of a screening assay. For example, to screen samples for the
20 presence of particular proteases (*e.g.*, proteases involved in blood clotting such as proteases Xa and VIIa), the samples can be assayed, for example using probes which are fluorogenic substrates specific for each protease of interest. If a target protease binds to and cleaves a substrate, the substrate will fluoresce, usually as a result, *e.g.*, of cleavage and separation between two energy transfer pairs, and the signal can be detected. In another example, to
25 screen samples for the presence of a particular kinase(s) (*e.g.*, a tyrosine kinase), samples containing one or more kinases of interest can be assayed, *e.g.*, using probes are peptides which can be selectively phosphorylated by one of the kinases of interest. Using art-recognized, routinely determinable conditions, samples can be incubated with an array of substrates, in an appropriate buffer and with the necessary cofactors, for an empirically
30 determined period of time. If necessary, reactions can be stopped, *e.g.*, by washing and the phosphorylated substrates can be detected by, for example, incubating them with detectable reagents such as, *e.g.*, fluorescein-labeled anti-phosphotyrosine or anti-phosphoserine antibodies and the signal can be detected.

In another embodiment, the compositions of the invention can be used to screen for agents which modulate a pattern of gene expression. Arrays of oligonucleotides can be used, for example, to identify mRNA species whose pattern of expression from a set of genes is correlated with a particular physiological state or developmental stage, or with a disease condition (“correlative” genes, RNAs, or expression patterns). By the terms “correlate” or “correlative,” it is meant that the synthesis pattern of RNA is associated with the physiological condition of a cell, but not necessarily that the expression of a given RNA is responsible for or is causative of a particular physiological state. For example, a small subset of mRNAs can be identified which are modulated (*e.g.*, upregulated or downregulated) in cells which serve as a model for a particular disease state. This altered pattern of expression as compared to that in a normal cell, which does not exhibit a pathological phenotype, can serve as a indicator of the disease state (“indicator” or “correlative” genes, RNAs, or expression patterns).

The invention also relates to a selecting oligonucleotides for the methods described herein in which many oligomers are screened (*e.g.*, from about 10-20 to significantly greater numbers as may be found in a combinatorial library), after which the more efficacious oligomers are chosen and combined to produce a composition of the invention. Thus, inhibition of greater than 95%, 90%, 85%, 80%, 70%, or 60% may be achieved.

Compositions which modulate the chosen indicator expression pattern (*e.g.*, compared to control compositions comprising, for example oligonucleotides which comprise a nucleotide sequence which is the reverse of the oligonucleotide, or which contains mismatch bases) can indicate that a particular target gene is a potential target for therapeutic intervention. Moreover, such compositions may be useful as therapeutic agents to modulate expression patterns of cells in an *in vitro* expression system or in *in vivo* therapy. As used herein, “modulate” means to cause to increase or decrease the amount or activity of a molecule or the like which is involved in a measurable reaction. In one embodiment, a series of cells (*e.g.*, from a disease model) can be contacted with a series of agents (*e.g.*, for a period of time ranging from about 10 minutes to about 48 hours or more) and, using routine, art-recognized methods (*e.g.*, commercially available kits), total RNA or mRNA extracts can be made. If it is desired to amplify the amount of RNA, standard procedures such as RT-PCR amplification can be used (see, *e.g.*, Innis *et al.* eds., (1996) PCR Protocols: A Guide to Methods in Amplification, Academic Press, New York). The extracts (or amplified products from them) can be allowed to contact (*e.g.*, incubate with) probes for appropriate indicator RNAs, and those agents which are associated with a change in the indicator expression pattern can be identified.

Similarly, agents can be identified which modulate expression patterns associated with particular physiological states or developmental stages. Such agents can be man-made or naturally-occurring substances, including environmental factors such as substances involved in embryonic development or in regulating physiological reactions.

- 5 In one embodiment, the methods described herein can be performed in a “high throughput” manner, in which a large number of target genes (*e.g.*, as many as about 1000 or more, depending on the particular format used) are assayed rapidly and concurrently. Further, many assay formats (*e.g.*, plates or surfaces) can be processed at one time. For example, because the oligonucleotides of the invention do not need to be tested individually
- 10 before incorporating them into a composition, they can be readily synthesized and large numbers of target genes can be tested at one time. For example, a large number of samples, each comprising a biological sample containing a target nucleic acid molecule (*e.g.*, a cell) and a composition of the invention can be added to separate regions of an assay format and assays can be performed on each of the samples.

15

Administration of Oligonucleotide Compositions

- The optimal course of administration or delivery of the oligonucleotides may vary depending upon the desired result and/ or on the subject to be treated. As used herein “administration” refers to contacting cells with oligonucleotides and can be performed *in*
- 20 *vitro* or *in vivo*. The dosage of oligonucleotides may be adjusted to optimally reduce expression of a protein translated from a target nucleic acid molecule, *e.g.*, as measured by a readout of RNA stability or by a therapeutic response, without undue experimentation.

- For example, expression of the protein encoded by the nucleic acid target can be measured to determine whether or not the dosage regimen needs to be adjusted accordingly.
- 25 In addition, an increase or decrease in RNA or protein levels in a cell or produced by a cell can be measured using any art recognized technique. By determining whether transcription has been decreased, the effectiveness of the oligonucleotide in inducing the cleavage of a target RNA can be determined.

- Any of the above-described oligonucleotide compositions can be used alone or in
- 30 conjunction with a pharmaceutically acceptable carrier. As used herein, “pharmaceutically acceptable carrier” includes appropriate solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like. The use of such media and agents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active ingredient, it can
- 35 be used in the therapeutic compositions. Supplementary active ingredients can also be incorporated into the compositions.

Oligonucleotides may be incorporated into liposomes or liposomes modified with polyethylene glycol or admixed with cationic lipids for parenteral administration. Incorporation of additional substances into the liposome, for example, antibodies reactive against membrane proteins found on specific target cells, can help target the oligonucleotides to specific cell types.

Moreover, the present invention provides for administering the subject oligonucleotides with an osmotic pump providing continuous infusion of such oligonucleotides, for example, as described in Rataiczak, *et al.* (1992 *Proc. Natl. Acad. Sci. USA* 89:11823-11827). Such osmotic pumps are commercially available, *e.g.*, from Alzet Inc. (Palo Alto, Calif.). Topical administration and parenteral administration in a cationic lipid carrier are preferred.

With respect to *in vivo* applications, the formulations of the present invention can be administered to a patient in a variety of forms adapted to the chosen route of administration, *e.g.*, parenterally, orally, or intraperitoneally. Parenteral administration, which is preferred, includes administration by the following routes: intravenous; intramuscular; interstitially; intraarterially; subcutaneous; intra ocular; intrasynovial; trans epithelial, including transdermal; pulmonary via inhalation; ophthalmic; sublingual and buccal; topically, including ophthalmic; dermal; ocular; rectal; and nasal inhalation via insufflation.

Pharmaceutical preparations for parenteral administration include aqueous solutions of the active compounds in water-soluble or water-dispersible form. In addition, suspensions of the active compounds as appropriate oily injection suspensions may be administered. Suitable lipophilic solvents or vehicles include fatty oils, for example, sesame oil, or synthetic fatty acid esters, for example, ethyl oleate or triglycerides. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension include, for example, sodium carboxymethyl cellulose, sorbitol, or dextran, optionally, the suspension may also contain stabilizers. The oligonucleotides of the invention can be formulated in liquid solutions, preferably in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the oligonucleotides may be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included in the invention.

Pharmaceutical preparations for topical administration include transdermal patches, ointments, lotions, creams, gels, drops, sprays, suppositories, liquids and powders. In addition, conventional pharmaceutical carriers, aqueous, powder or oily bases, or thickeners may be used in pharmaceutical preparations for topical administration.

Pharmaceutical preparations for oral administration include powders or granules, suspensions or solutions in water or non-aqueous media, capsules, sachets or tablets. In addition, thickeners, flavoring agents, diluents, emulsifiers, dispersing aids, or binders may be used in pharmaceutical preparations for oral administration.

5 For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives, and detergents. Transmucosal administration may be through nasal sprays or using suppositories. For oral administration, the oligonucleotides are formulated into conventional
10 oral administration forms such as capsules, tablets, and tonics. For topical administration, the oligonucleotides of the invention are formulated into ointments, salves, gels, or creams as known in the art.

Drug delivery vehicles can be chosen *e.g.*, for *in vitro*, for systemic, or for topical administration. These vehicles can be designed to serve as a slow release reservoir or to
15 deliver their contents directly to the target cell. An advantage of using some direct delivery drug vehicles is that multiple molecules are delivered per uptake. Such vehicles have been shown to increase the circulation half-life of drugs that would otherwise be rapidly cleared from the blood stream. Some examples of such specialized drug delivery vehicles which fall into this category are liposomes, hydrogels, cyclodextrins, biodegradable nanocapsules, and
20 bioadhesive microspheres.

The described oligonucleotides may be administered systemically to a subject. Systemic absorption refers to the entry of drugs into the blood stream followed by distribution throughout the entire body. Administration routes which lead to systemic absorption include: intravenous, subcutaneous, intraperitoneal, and intranasal. Each of these
25 administration routes delivers the oligonucleotide to accessible diseased cells. Following subcutaneous administration, the therapeutic agent drains into local lymph nodes and proceeds through the lymphatic network into the circulation. The rate of entry into the circulation has been shown to be a function of molecular weight or size. The use of a liposome or other drug carrier localizes the oligonucleotide at the lymph node. The
30 oligonucleotide can be modified to diffuse into the cell, or the liposome can directly participate in the delivery of either the unmodified or modified oligonucleotide into the cell.

The chosen method of delivery will result in entry into cells. Preferred delivery methods include liposomes (10-400 nm), hydrogels, controlled-release polymers, and other pharmaceutically applicable vehicles, and microinjection or electroporation (for *ex vivo*
35 treatments).

The pharmaceutical preparations of the present invention may be prepared and formulated as emulsions. Emulsions are usually heterogenous systems of one liquid dispersed in another in the form of droplets usually exceeding 0.1 μm in diameter.

5 The emulsions of the present invention may contain excipients such as emulsifiers, stabilizers, dyes, fats, oils, waxes, fatty acids, fatty alcohols, fatty esters, humectants, hydrophilic colloids, preservatives, and anti-oxidants may also be present in emulsions as needed. These excipients may be present as a solution in either the aqueous phase, oily phase or itself as a separate phase.

10 Examples of naturally occurring emulsifiers that may be used in emulsion formulations of the present invention include lanolin, beeswax, phosphatides, lecithin and acacia. Finely divided solids have also been used as good emulsifiers especially in combination with surfactants and in viscous preparations. Examples of finely divided solids that may be used as emulsifiers include polar inorganic solids, such as heavy metal hydroxides, nonswelling clays such as bentonite, attapulgite, hectorite, kaolin,
15 montmorillonite, colloidal aluminum silicate and colloidal magnesium aluminum silicate, pigments and nonpolar solids such as carbon or glyceryl tristearate.

Examples of preservatives that may be included in the emulsion formulations include methyl paraben, propyl paraben, quaternary ammonium salts, benzalkonium chloride, esters of p-hydroxybenzoic acid, and boric acid. Examples of antioxidants that may be included in
20 the emulsion formulations include free radical scavengers such as tocopherols, alkyl gallates, butylated hydroxyanisole, butylated hydroxytoluene, or reducing agents such as ascorbic acid and sodium metabisulfite, and antioxidant synergists such as citric acid, tartaric acid, and lecithin.

In one embodiment, the compositions of oligonucleotides are formulated as
25 microemulsions. A microemulsion is a system of water, oil and amphiphile which is a single optically isotropic and thermodynamically stable liquid solution. Typically microemulsions are prepared by first dispersing an oil in an aqueous surfactant solution and then adding a sufficient amount of a 4th component, generally an intermediate chain-length alcohol to form a transparent system.

Surfactants that may be used in the preparation of microemulsions include, but are not limited to, ionic surfactants, non-ionic surfactants, Brij 96, polyoxyethylene oleyl ethers, polyglycerol fatty acid esters, tetraglycerol monolaurate (ML310), tetraglycerol monooleate (MO310), hexaglycerol monooleate (PO310), hexaglycerol pentaoleate (PO500),
5 decaglycerol monocaprate (MCA750), decaglycerol monooleate (MO750), decaglycerol sequioleate (S0750), decaglycerol decaoleate (DA0750), alone or in combination with cosurfactants. The cosurfactant, usually a short-chain alcohol such as ethanol, 1-propanol, and 1-butanol, serves to increase the interfacial fluidity by penetrating into the surfactant film and consequently creating a disordered film because of the void space generated among surfactant
10 molecules.

Microemulsions may, however, be prepared without the use of cosurfactants and alcohol-free self-emulsifying microemulsion systems are known in the art. The aqueous phase may typically be, but is not limited to, water, an aqueous solution of the drug, glycerol, PEG300, PEG400, polyglycerols, propylene glycols, and derivatives of ethylene glycol. The
15 oil phase may include, but is not limited to, materials such as Captex 300, Captex 355, Capmul MCM, fatty acid esters, medium chain (C8-C12) mono, di, and tri-glycerides, polyoxyethylated glyceryl fatty acid esters, fatty alcohols, polyglycolized glycerides, saturated polyglycolized C8-C10 glycerides, vegetable oils and silicone oil.

Microemulsions are particularly of interest from the standpoint of drug solubilization
20 and the enhanced absorption of drugs. Lipid based microemulsions (both oil/water and water/oil) have been proposed to enhance the oral bioavailability of drugs.

Microemulsions offer improved drug solubilization, protection of drug from enzymatic hydrolysis, possible enhancement of drug absorption due to surfactant-induced alterations in membrane fluidity and permeability, ease of preparation, ease of oral
25 administration over solid dosage forms, improved clinical potency, and decreased toxicity (Constantinides *et al.*, *Pharmaceutical Research*, 1994, 11:1385; Ho *et al.*, *J. Pharm. Sci.*, 1996, 85:138-143). Microemulsions have also been effective in the transdermal delivery of active components in both cosmetic and pharmaceutical applications. It is expected that the microemulsion compositions and formulations of the present invention will facilitate the
30 increased systemic absorption of oligonucleotides from the gastrointestinal tract, as well as improve the local cellular uptake of oligonucleotides within the gastrointestinal tract, vagina, buccal cavity and other areas of administration.

In an embodiment, the present invention employs various penetration enhancers to effect the efficient delivery of nucleic acids, particularly oligonucleotides, to the skin of animals. Even non-lipophilic drugs may cross cell membranes if the membrane to be crossed is treated with a penetration enhancer. In addition to increasing the diffusion of non-lipophilic drugs across cell membranes, penetration enhancers also act to enhance the permeability of lipophilic drugs.

Five categories of penetration enhancers that may be used in the present invention include: surfactants, fatty acids, bile salts, chelating agents, and non-chelating non-surfactants. Other agents may be utilized to enhance the penetration of the administered oligonucleotides include: glycols such as ethylene glycol and propylene glycol, pyrrols such as 2-pyrrol, azones, and terpenes such as limonene and menthone.

The oligonucleotides, especially in lipid formulations, can also be administered by coating a medical device, for example, a catheter, such as an angioplasty balloon catheter, with a cationic lipid formulation. Coating may be achieved, for example, by dipping the medical device into a lipid formulation or a mixture of a lipid formulation and a suitable solvent, for example, an aqueous-based buffer, an aqueous solvent, ethanol, methylene chloride, chloroform and the like. An amount of the formulation will naturally adhere to the surface of the device which is subsequently administered to a patient, as appropriate. Alternatively, a lyophilized mixture of a lipid formulation may be specifically bound to the surface of the device. Such binding techniques are described, for example, in K. Ishihara *et al.*, Journal of Biomedical Materials Research, Vol. 27, pp. 1309-1314 (1993), the disclosures of which are incorporated herein by reference in their entirety.

The useful dosage to be administered and the particular mode of administration will vary depending upon such factors as the cell type, or for *in vivo* use, the age, weight and the particular animal and region thereof to be treated, the particular oligonucleotide and delivery method used, the therapeutic or diagnostic use contemplated, and the form of the formulation, for example, suspension, emulsion, micelle or liposome, as will be readily apparent to those skilled in the art. Typically, dosage is administered at lower levels and increased until the desired effect is achieved. When lipids are used to deliver the oligonucleotides, the amount of lipid compound that is administered can vary and generally depends upon the amount of oligonucleotide agent being administered. For example, the weight ratio of lipid compound to oligonucleotide agent is preferably from about 1:1 to about 15:1, with a weight ratio of about 5:1 to about 10:1 being more preferred. Generally, the amount of cationic lipid compound which is administered will vary from between about 0.1 milligram (mg) to about 1 gram (g). By way of general guidance, typically between about 0.1 mg and about 10 mg of the particular oligonucleotide agent, and about 1 mg to about 100 mg of the lipid compositions, each per kilogram of patient body weight, is administered, although higher and lower amounts can be used.

The agents of the invention are administered to subjects or contacted with cells in a biologically compatible form suitable for pharmaceutical administration. By "biologically compatible form suitable for administration" is meant that the oligonucleotide is administered in a form in which any toxic effects are outweighed by the therapeutic effects of the oligonucleotide. In one embodiment, oligonucleotides can be administered to subjects. Examples of subjects include mammals, *e.g.*, humans, cows, pigs, horses, dogs, cats, mice, rats, and transgenic non-human animals.

Administration of an active amount of an oligonucleotide of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, an active amount of an oligonucleotide may vary according to factors such as the type of cell, the oligonucleotide used, and for *in vivo* uses the disease state, age, sex, and weight of the individual, and the ability of the oligonucleotide to elicit a desired response in the individual. Establishment of therapeutic levels of oligonucleotides within the cell is dependent upon the rates of uptake and efflux or degradation. Decreasing the degree of degradation prolongs the intracellular half-life of the oligonucleotide. Thus, chemically-modified oligonucleotides, *e.g.*, with modification of the phosphate backbone, may require different dosing.

The exact dosage of an oligonucleotide and number of doses administered will depend upon the data generated experimentally and in clinical trials. Several factors such as the desired effect, the delivery vehicle, disease indication, and the route of administration, will affect the dosage. Dosages can be readily determined by one of ordinary skill in the art and formulated into the subject pharmaceutical compositions. Preferably, the duration of treatment will extend at least through the course of the disease symptoms.

Dosage regima may be adjusted to provide the optimum therapeutic response. For example, the oligonucleotide may be repeatedly administered, *e.g.*, several doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation. One of ordinary skill in the art will readily be able to determine appropriate doses and schedules of administration of the subject oligonucleotides, whether the oligonucleotides are to be administered to cells or to subjects.

Treatment of Diseases or Disorders

By inhibiting the expression of a gene, the oligonucleotide compositions of the present invention can be used to treat any disease involving the expression of a protein. Examples of diseases that can be treated by oligonucleotide compositions include: cancer, retinopathies, autoimmune diseases, inflammatory diseases (*e.g.*, ICAM-1 related disorders, Psoriasis, Ulcerative Colitus, Crohn's disease), viral diseases (*e.g.*, HIV, Hepatitis C), and cardiovascular diseases.

In one embodiment, *in vitro* treatment of cells with oligonucleotides can be used for *ex vivo* therapy of cells removed from a subject (*e.g.*, for treatment of leukemia or viral infection) or for treatment of cells which did not originate in the subject, but are to be administered to the subject (*e.g.*, to eliminate transplantation antigen expression on cells to be transplanted into a subject). In addition, *in vitro* treatment of cells can be used in non-therapeutic settings, *e.g.*, to evaluate gene function, to study gene regulation and protein synthesis or to evaluate improvements made to oligonucleotides designed to modulate gene expression or protein synthesis. *In vivo* treatment of cells can be useful in certain clinical settings where it is desirable to inhibit the expression of a protein. There are numerous medical conditions for which antisense therapy is reported to be suitable (see *e.g.*, U.S. patent 5,830,653) as well as respiratory syncytial virus infection (WO 95/22553) influenza virus (WO 94/23028), and malignancies (WO 94/08003). Other examples of clinical uses of antisense oligonucleotides are reviewed, *e.g.*, in Glaser. 1996. *Genetic Engineering News* 16:1. Exemplary targets for cleavage by oligonucleotides include *e.g.*, protein kinase Ca, ICAM-1, c-raf kinase, p53, c-myb, and the bcr/abl fusion gene found in chronic myelogenous leukemia.

- The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques
- 5 are explained fully in the literature. See, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, J. *et al.* (Cold Spring Harbor Laboratory Press (1989)); *Short Protocols in Molecular Biology*, 3rd Ed., ed. by Ausubel, F. *et al.* (Wiley, NY (1995)); *DNA Cloning*, Volumes I and II (D. N. Glover ed., 1985); *Oligonucleotide Synthesis* (M. J. Gait ed. (1984)); Mullis *et al.* U.S. Patent No: 4,683,195; *Nucleic Acid Hybridization* (B. D.
- 10 Hames & S. J. Higgins eds. (1984)); the treatise, *Methods In Enzymology* (Academic Press, Inc., N.Y.); *Immunochemical Methods In Cell And Molecular Biology* (Mayer and Walker, eds., Academic Press, London (1987)); *Handbook Of Experimental Immunology*, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds. (1986)); and Miller, J. *Experiments in Molecular Genetics* (Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1972)).
- 15 The invention is further illustrated by the following examples, which should not be construed as further limiting.

EXAMPLES

20 Example 1. Ability of Oligonucleotide Compositions to Inhibit CDK2 in A549 Cells.

- In this example, the ability of 5 different antisense oligonucleotides individually was compared with the ability of all 5 of the antisense oligonucleotides transfected at one time for their ability to inhibit the expression of CDK2 in A549 cells. The sequences of the 5 antisense oligonucleotides used were: Oligonucleotide 1 GCAGUAUACCUCUCGCU-
- 25 CUUGUCAA (SEQ ID NO: ##); oligonucleotide 2 UUUGGAAGUUCUCCAUGAA-GCGCCA (SEQ ID NO: ##); oligonucleotide 3 GUCCAAAGUCUGCUA-GCUUGAUGGC (SEQ ID NO: ##); oligonucleotide 4 CCCAGGAGGAUUU-CAGGAGCUCGGU (SEQ ID NO: ##); oligonucleotide 5 UAGAAGUAACUCCU-GGCCACACCAC (SEQ ID NO: ##); reverse control AACUGUUCUCGCUC-
- 30 UCCAUAUGACG (SEQ ID NO: ##).

For transfection with antisense oligonucleotides A549 cells were maintained in DMEM with high glucose (Gibco-BRL) supplemented with 10% Fetal Bovine Serum, 2mM L-Glutamine, and 1X penicillin/streptomycin.

On the day before transfection 24-well plates were seeded with 30,000 A549 cells per well. The cells were approximately 60% confluent at the start of transfection, and were evenly distributed across the plate. On the day of transfection, a 10X stock of Lipofectamine 2000 (Invitrogen) was prepared in Opti-MEM (serum free media, Gibco-BRL). The diluted lipid was allowed to stand at room temperature for 15 minutes. The optimal conditions for transfection of A549 cells were determined to be 25 nM oligonucleotide complexed with 1 ug/mL Lipofectamine 2000. A 10X stock of each oligonucleotide to be used in the transfection was also prepared in Opti-MEM (10X concentration of oligonucleotide is 0.25 uM). Equal volumes of the 10X Lipofectamine 2000 stock and the 10X oligonucleotide solutions were mixed well and incubated for 15 minutes at room temperature to allow complexation of the oligonucleotide and lipid. The resulting mixture was 5X. After the 15 minutes of complexation, four volumes of full growth media was added to the oligonucleotide/lipid complexes to make a 1X solution. The media was aspirated from the cells, and 0.5 mL of the 1X oligonucleotide/lipid complexes was added to each well. The cells were not permitted to dry out during the changing of media. The cells were incubated for 16-24 hours at 37°C in a humidified CO₂ incubator. Cell pellets were harvested for protein determination or RNA isolation. The Tables below show the results of the experiment.

Oligonucleotide	Ratio of CDK2 expression to GAPDH expression	Standard Deviation
No transfection	1.481	0.242
FITC	1.004	0.203
1	0.233	0.041
2	0.231	0.058
3	0.198	0.015
4	0.193	0.065
5	0.673	0.232
Reverse Control	0.749	0.079
Oligonucleotide Composition	0.137	0.012

20

Oligonucleotide	Percent Inhibition Compared to Reverse Control
No transfection	0 (-98%)
FITC	0 (-34%)
1	69%

Oligonucleotide	Percent Inhibition Compared to Reverse Control
2	69%
3	74%
4	74%
5	10%
Reverse Control	0%
Oligonucleotide Composition	82%

The levels of expression of CDK2 were normalized to levels of GAPDH. No transfection or transfection with a fluorescent control oligonucleotide (which targets luciferase) showed levels of 1 or higher. A reverse sequence control oligonucleotide gave a level of about 0.8. Each of the individual oligonucleotides (1-5) showed inhibition in CDK2 expression (with levels ranging from about 0.2 (about 70% inhibition compared to the reverse control) to 0.65 (10% inhibition compared to the reverse control) for oligonucleotide number 5). All five of the oligonucleotides transfected at once gave a level of less than about 0.2, about 82% inhibition compared to the reverse control. Thus, using only one transfection, an oligonucleotide composition comprising five different antisense oligonucleotides can be used to efficiently inhibit protein synthesis.

Example 2. Summary of Results of Experiments in Which Oligonucleotide Compositions Were Tested on Thirty Different Genes.

Figure 1 shows a summary of the results of about 30 antisense inhibition experiments against about thirty different genes in cell culture. Antisense was transfected as described in Example 1 and inhibition analyzed by Taqman real time PCR using standard methods. In each case the antisense inhibition was determined by comparison to a control oligonucleotide of the same chemistry that was not antisense to the target gene. Antisense compositions comprised 5-8 antisense oligonucleotides that had been designed against each gene, and individual oligonucleotides where compared to the mixtures of 5 or more antisense oligonucleotides. For three target genes the mixtures did not work well, and these data were eliminated from the analysis of the mixtures. Remarkably, the mixtures inhibited approximately as well (81-vs 84%) as the best individual oligonucleotide. The average inhibition of all individual oligonucleotides was much lower (56%), with a much higher variation. Thus, using the mixtures allows one to obtain high inhibition in the vast majority of cases (~90% of the target genes) without first screening through individual oligonucleotides to select those which work best. Also, as evidenced by the increased variation in the results obtained when individual oligonucleotides were used, in many cases the mixture was better than the best individual oligonucleotide.

Example 3. Ultramer data for a mixture of siRNA complexes targeting p53.

HeLa cells were transfected with 50 nM siRNA complexed with 1 ug/mL of Lipofectamine 2000 for 24 hours. After 24 hours, cells were lysed and RNA isolated for analysis by RT-PCR. Seven siRNA complexes were transfected that target a unique site of the p53 gene and a mixture of all seven siRNAs (equal concentrations of each) called the "siRNA ultramer." The best siRNA complex inhibited the target by 87% and the ultramer inhibited 69% compared to average of the controls.

P53 sequences (Antisense, Sense):

siRNA1:

CUGACUGCGGCUCCUCCAUTT (SEQ ID NO: ##)
AUGGAGGAGCCGCAGUCAGTT (SEQ ID NO: ##)

siRNA2:

CUCACAACCUCCGUCAUGUTT (SEQ ID NO: ##)
ACAUGACGGAGGUUGUGAGTT (SEQ ID NO: ##)

siRNA3:

GACCAUCGCUAUCUGAGCATT (SEQ ID NO: ##)

UGCUCAGAUAGCGAUGGUCTT (SEQ ID NO: ##)
 siRNA4:
 GUACAGUCAGAGCCAACCUTT (SEQ ID NO: ##)
 AGGUUGGCUCUGACUGUACTT (SEQ ID NO: ##)
 5 siRNA5:
 ACCUCAAGCUGUCCGUCTT (SEQ ID NO: ##)
 GACGGAACAGCUUGAGGUTT (SEQ ID NO: ##)
 siRNA6:
 CCUCAUUCAGCUCUCGGAATT (SEQ ID NO: ##)
 10 UUCCGAGAGCUGAAUGAGGTT (SEQ ID NO: ##)
 siRNA7:
 CCCUUCUGUCUUGAACAUGTT (SEQ ID NO: ##)
 CAUGUUCAAGACAGAAGGGTT (SEQ ID NO: ##)

15 Example 4. Ultramer data for a mixture of siRNA complexes targeting GTP20.

Human Mesenchymal Stems cells (hMSC) were transfected with 2 ug/mL
 Lipofectamine 2000 complexed to 400 nM siRNA (total concentration, for clarity in the
 mixture each individual oligomer was at 80nM). Five siRNA duplexes targeted to GTP20
 (TD), one composition matched control duplex (CD) and an equimolar mixture of each of the
 20 5 oligos ("Mixture") were transfected continuously for 24 hours and RNA was harvested
 using the RNA Catcher (Sequitur, Inc. Natick, MA). Expression of GTP20 mRNA was
 quantified by Taqman and normalized to GAPDH. Inhibition of 70% or greater relative to
 the control duplex was achieved using TD5 (70%) and the Ultramer (76%).

Human mesenchymal stem cells were plated at 15,000 per well in 48 well dishes and
 25 transfected 24 hours later. Lipofectamine 2000 was diluted in Opti-MEM to a 10X
 concentration of 20 ug/mL and incubated for 15 minutes. Following incubation, lipid was
 complexed to siRNA duplexes by addition of 10X lipid to an equal volume of 10X (4uM)
 siRNA, and incubated for 15 minutes. 5X lipid/siRNA complexes were diluted to 1X by the
 addition of MSC Differentiation Media. 250 ul of each 1X siRNA treatment was added per
 30 well of 48 well dish. Each treatment was applied to triplicate wells. Osteoblastic
 differentiation of MSC was induced approximately 4 hours after transfection. Cells were
 differentiated for 4 days prior to RNA isolation.

Example 5. Ultraspeed data for a mixture of siRNA complexes targeting Cbfa-1.

- Human Mesenchymal Stems cells (hMSC) were transfected with 2 ug/mL Lipofectamine 2000 complexed to 400 nM siRNA (total concentration, in mixture each individual duplex was at 80nM). Five targeted duplexes (TD), five control duplexes (CD), one equimolar mixture of all 5 duplexes ("Mixture") and one control Mixture(UC) were transfected continuously for 72 hours. RNA was harvested 96 hours after transfection using the RNA Catcher. Expression of Cbfa-1 mRNA was quantified by Taqman and normalized to GAPDH. Inhibition of 70% or greater relative to the average of the control duplexes was achieved using TD4 (74%). The Mixture inhibited 70% relative to the Mixture Control.
- Human mesenchymal stem cells were plated at 15,000 per well in 48 well dishes and transfected 24 hours later. Lipofectamine 2000 was diluted in Opti-MEM to a 10X concentration of 20 ug/mL and incubated for 15 minutes. Following incubation, lipid was complexed to siRNA duplexes by addition of 10X lipid to an equal volume of 10X (4uM) siRNA, and incubated for 15 minutes. 5X lipid/siRNA complexes were diluted to 1X by the addition of MSC Differentiation Media. 250 ul of each 1X siRNA treatment was added per well of 48 well dish. Each treatment was applied to triplicate wells. Osteoblastic differentiation of MSC was induced approximately 4 hours after transfection. Cells were differentiated for 4 days prior to RNA isolation. The following antisense sequences of Cbfa-1 siRNA duplexes were used (corresponding sense sequences where the complementary sequence with a 2nt TT 3' overhang, T's are DNA, all other nucleotides are RNA):

	TD1 (s18883) : AUUUAAGCGUGCUGCCATT (SEQ ID NO: ##)
	TD2 (s18885) : CUGUAAUCUGACUCUGUCCTT (SEQ ID NO: ##)
	TD3 (s18887) : AAUAUGGUCGCCAAACAGATT (SEQ ID NO: ##)
25	TD4 (s18889) : GUCAACACCAUCAUUCUGGTT (SEQ ID NO: ##)
	TD5 (s18891) : AGGUUUAGAGUCAUCAAGCTT (SEQ ID NO: ##)
	CD1 (s18884) : ACCGUCGUGCGAUAAUUUATT (SEQ ID NO: ##)
	CD2 (s18886) : CCUGUCUCAGUCUAAUGUCTT (SEQ ID NO: ##)
	CD3 (s18888) : AGACAAACCGCUGGUUAUAATT (SEQ ID NO: ##)
30	CD4 (s18890) : GGUCUUACUACCACAACUGTT (SEQ ID NO: ##)
	CD5 (s18892) : CGAACUACUGAGAUUUGGATT (SEQ ID NO: ##)

Example 6. Ultramer data for a mixture of siRNA complexes targeting PTP mu.

Efficacy of all phosphorothioate DNA 25nt antisense oligonucleotides targeted against PTP mu mRNA in human lung carcinoma (A549) cells. Potent inhibition of mRNA was obtained following a 16 hour transfection of A549 cells with 25 nM oligo. AS: antisense oligonucleotide; RC: reverse control; MIX: mixture of individual AS oligomers (total oligomer concentration of 25 nM). Target mRNA quantity was normalized to GAPDH.

A549 cells at passage 3 were plated at 25,000 cells/well in 48 well plates and incubated overnight in a humidified 5% CO₂ chamber (37 °C). A 250 nM solution of AS oligomer in Optimem-I (Gibco BRL) was mixed with an equal volume of 10 ug/mL lipofectamine 2000 (Invitrogen) in Optimem-I (lipid solution was pre-incubated at 25C for 15 minutes). Oligomer-lipid complexes were formed by incubation at room temperature for 15 minutes. 4 volumes of DMEM plus 10% fetal serum medium was added to the complexes and 250 ul of the diluted suspension was added to cells. The final concentration of oligomer was 25 nM. Following a 16h transfection, cells were washed with PBS and poly A+ mRNA was isolated using Sequitur's mRNA Catcher. mRNA was quantified by real time RT-PCR (Taqman); automated data collection was with an ABI prism® sequence detection system. Data are normalized to GAPDH mRNA. Oligonucleotide sequences: AS1, CAUUCA-CCAGCAUGAGAGAACCUGA (SEQ ID NO: ##); AS2, TCCCAGAGGCAT-TCACCAGCATGAG (SEQ ID NO: ##); AS3, UCCAGAUAGGAUUCCC-CAGUGGCCC (SEQ ID NO: ##); AS4, CUGGUCAGGAGCACACUAAUCUCAU (SEQ ID NO: ##); AS5, AGUCAAGGUGUUCACUUGCUCCCAA (SEQ ID NO: ##); AS6, AAGUACUAAUGGCCAGUUCUGCCC (SEQ ID NO: ##); AS7, CCCUGUAACCAGAGCCUGUCUCCUG (SEQ ID NO: ##); AS8, GAGCUGG-UCACCUUGAUUCCUUCA (SEQ ID NO: ##); AS9, CCAGGCAAGUCCCAAGU-GUCCUCAU (SEQ ID NO: ##); AS10, GAUGUCCUAAACACCUUCACCUCAUC (SEQ ID NO: ##); MIX, equimolar solution of AS1 through AS10.

Example 7. Ultramer data for a mixture of siRNA complexes targeting PTP-PEST.

Efficacy of 25nt phosphorothioate DNA antisense oligonucleotides targeted against PTP-PEST mRNA in Human Umbilical Vein Endothelial Cells (HuVEC). Inhibition of mRNA was obtained following a 4 hour serum- free transfection of cells with 200 nM oligo followed by a 14 h incubation in serum-containing medium. AS: antisense oligonucleotide; RC: reverse control; Mixture: mixture of individual AS oligomers (total oligo concentration of 200 nM). Target mRNA quantity is normalized to GAPDH.

HuVEC cells at passage 3 were plated at 25,000 cells/well in 48 well plates and incubated overnight in a humidified 5% CO₂ chamber (37°C). A 2000 nM solution of AS oligomer in Optimem-I (Gibco BRL) was mixed with an equal volume of 100 ug/mL Lipofectin (Gibco BRL) in Optimem-I (lipid solution was pre-incubated at 25°C for 30 minutes). Oligomer-lipid complexes were formed by incubation at room temperature for 30 minutes. 4 volumes of Optimem-I (serum-free) was added to the complexes and 250 ul of the diluted suspension was added to cells. Four hours later, the transfection complexes were aspirated and replaced with 250 ul of EGM-2 complete serum medium (Clonetics/Biowhittaker). Following a 16h transfection, cells were washed with PBS and poly A+ mRNA was isolated using an mRNA Catcher (Sequitur, Inc.). mRNA was quantified by real time RT-PCR (Taqman); automated data collection was with an ABI prism® sequence detection system.

Data are normalized to GAPDH mRNA. AS1, CCCAUUGUGGUCAGGAC-UCUUCAUGU (SEQ ID NO: ##); AS2, UUCCCAUCUCAAAUUCU-CGGCAGGCU (SEQ ID NO: ##); AS3, UGGCACAAAUGGCACCUGUUCUCCU (SEQ ID NO: ##); RC, GACUCCUUUAAGUAGGUCUCCAGGU (SEQ ID NO: ##). MIX, equimolar solution of AS1, AS2, and AS3.

Example 8. Ultramer data for a mixture of siRNA complexes targeting PTP-eta.

Efficacy of all phosphorothioate DNA 25nt antisense oligonucleotides targeted against PTP-eta mRNA in Normal Rat Kidney (NRK) cells. Inhibition of mRNA was obtained following an overnight transfection of cells with 25 nM oligo. AS: antisense oligonucleotide; RC: reverse control; Mix: mixture of individual AS oligomers (total oligomer concentration of 25 nM). Target mRNA quantity is normalized to GAPDH.

NRK cells at passage 5 were plated at 25,000 cells/well in 48 well plates and incubated overnight in a humidified 5% CO₂ chamber (37°C). A 250 nM solution of AS oligomer in Optimem-I (Gibco BRL) was mixed with an equal volume of 10 ug/mL Lipofectamine 2000 (Invitrogen) in Optimem-I (lipid solution was pre-incubated at 25C for 30 minutes). Oligomer-lipid complexes were formed by incubation at room temperature for 15 minutes. 4 volumes of complete DMEM plus 5% bovine calf serum were added to the complexes and 250 ul of the diluted suspension was layered onto cells. The final oligomer concentration was 25 nM. Following a 16h incubation, cells were washed with PBS and poly A+ mRNA was isolated using Sequitur's mRNA Catcher*. mRNA was quantified by real time RT-PCR (Taqman*); automated data collection was with an ABI prism® sequence detection system.

Data are normalized to GAPDH mRNA. AS1, ACCUGUGCACACAACCUGGC-
CCUGGU (SEQ ID NO: ##); AS2, ACAGUAUACCGCAGCGUGUUUCCCUU
(SEQ ID NO: ##); AS3, GUCUCAUUGACUGUCCCAAGGUGAU (SEQ ID NO:
##); AS4, GCUCUACAAUCUGCAUCCGGUAAGAU (SEQ ID NO: ##); AS5,
5 UCUGUGCCAUCUGCUGCUUGAGAAUU (SEQ ID NO: ##); AS6, UGUUCAC-
AGCUCGGAUGUCAGAAACU (SEQ ID NO: ##); RC, UAAGAGUUCGUCGU-
CUACCGUGUCUU (SEQ ID NO: ##); MIX, equimolar solution of AS1 through AS6

Equivalents

- 10 Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

1. An oligonucleotide composition comprising at least 3 different oligonucleotides
5 targeted to at least three different nucleotide sequences within a target gene, wherein
(i) the oligonucleotides bind to their target nucleotide sequence with high affinity and
(ii) the oligonucleotides are GC enriched.
2. The oligonucleotide composition of claim 1, wherein the oligonucleotides are
10 antisense oligonucleotides.
3. The oligonucleotide composition of claim 1, wherein the oligonucleotides are double-
stranded RNA oligonucleotides.
- 15 4. The oligonucleotide composition of claim 1, wherein the oligonucleotide
compositions bind to their target nucleotide sequence with a T_m of at least about
60°C.
5. The oligonucleotide composition of claim 1, wherein the oligonucleotides have a GC
20 content of at least about 20%.
6. The oligonucleotide composition of claim 1, wherein the composition comprises at
least about 4 antisense oligonucleotides targeting at least four different nucleic acid
sequences.
- 25 7. The oligonucleotide composition of claim 1, wherein the composition comprises at
least about 5 oligonucleotides targeting at least five different nucleic acid sequences.
8. The oligonucleotide composition of claim 1, wherein the composition comprises at
30 least about 6 oligonucleotides targeting at least six different nucleic acid sequences.

9. The oligonucleotide composition of claim 1, wherein the oligonucleotides are at least about 25 nucleomonomers in length.
- 5 10. The oligonucleotide composition of claim 1, wherein the oligonucleotides are greater than 25 nucleomonomers in length.
11. The oligonucleotide composition of claim 2, wherein at least one of the antisense oligonucleotides is complementary in sequence to its target nucleotide sequence.
- 10 12. The oligonucleotide composition of claim 2, wherein the antisense oligonucleotides activate RNase H.
13. The oligonucleotide composition of claim 1, wherein at least one of the oligonucleotides comprise at least one modified internucleoside linkage.
- 15 14. The oligonucleotide composition of claim 1, wherein at least one of the oligonucleotides comprise at least one modified sugar moiety.
15. The oligonucleotide composition of claim 1, further comprising a pharmaceutically acceptable carrier.
- 20 16. The oligonucleotide composition of claim 1, wherein the oligonucleotide composition achieves a level of inhibition of protein synthesis the same as or higher than the level of inhibition achieved by the most effective individual oligonucleotide of the composition.
- 25 17. The oligonucleotide composition of claim 1, wherein the individual oligonucleotides are not separately tested for their ability to inhibit protein synthesis prior to their incorporation into the composition.
- 30 18. The oligonucleotide composition of claim 1, wherein the oligonucleotide composition results in greater than about 80% inhibition of protein synthesis.

19. A method of inhibiting protein synthesis in a cell comprising contacting the cell with at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched, to thereby inhibit protein synthesis.
20. The method of claim 19, wherein the oligonucleotides are antisense oligonucleotides.
21. The method of claim 19, wherein the oligonucleotides are double-stranded RNA oligonucleotides.
22. The method of claim 19, wherein the method is performed in a high-throughput format.
23. A method of identifying function of a gene encoding a protein comprising: contacting the cell with at least 3 different oligonucleotides targeted to at least three different nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC enriched, and assaying for a change in a detectable phenotype in the cell resulting from the inhibition of protein expression, to thereby determine the function of a gene.
24. The method of claim 23, wherein the oligonucleotides are antisense oligonucleotides.
25. The method of claim 23, wherein the oligonucleotides are double-stranded RNA oligonucleotides.
26. The method of claim 23, wherein the method is performed in a high-throughput format.

27. A method of making the oligonucleotide composition of claim 1, comprising:
combining at least 3 different oligonucleotides targeted to at least three different
nucleotide sequences within a target gene, wherein (i) the oligonucleotides bind to
their target nucleotide sequence with high affinity and (ii) the oligonucleotides are GC
enriched, and wherein, the individual oligonucleotides are not separately tested for
their ability to inhibit protein synthesis prior to their incorporation into the
composition.
28. The method of claim 27, wherein the oligonucleotides are antisense oligonucleotides.
29. The method of claim 27, wherein the oligonucleotides are double-stranded RNA
oligonucleotides.
30. An oligonucleotide composition comprising at least 3 different double-stranded RNA
oligonucleotides targeted to at least three different nucleotide sequences within a
target gene.
31. A method of inhibiting protein synthesis in a cell comprising contacting the cell with
at least 3 different double-stranded RNA oligonucleotides targeted to at least three
different nucleotide sequences within a target gene.
32. A method of identifying function of a gene encoding a protein comprising: contacting
the cell with at least 3 different double-stranded RNA oligonucleotides targeted to at
least three different nucleotide sequences within a target gene and assaying for a
change in a detectable phenotype in the cell resulting from the inhibition of protein
expression, to thereby determine the function of a gene.
33. A method of making an oligonucleotide composition comprising: combining at least 3
different double-stranded RNA oligonucleotides targeted to at least three different
nucleotide sequences within a target gene wherein, the individual oligonucleotides are
not separately tested for their ability to inhibit protein synthesis prior to their
incorporation into the composition.

FIGURE 1.

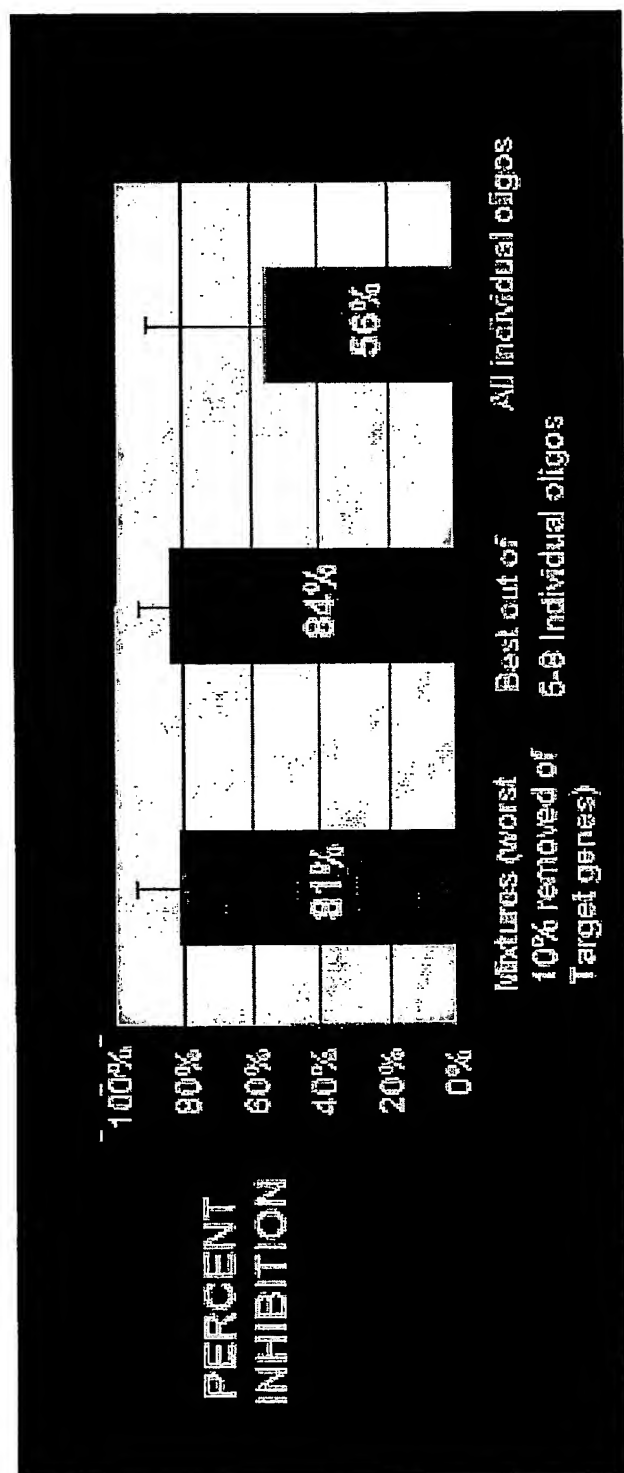


FIGURE 2.
siRNA Ultramer Data

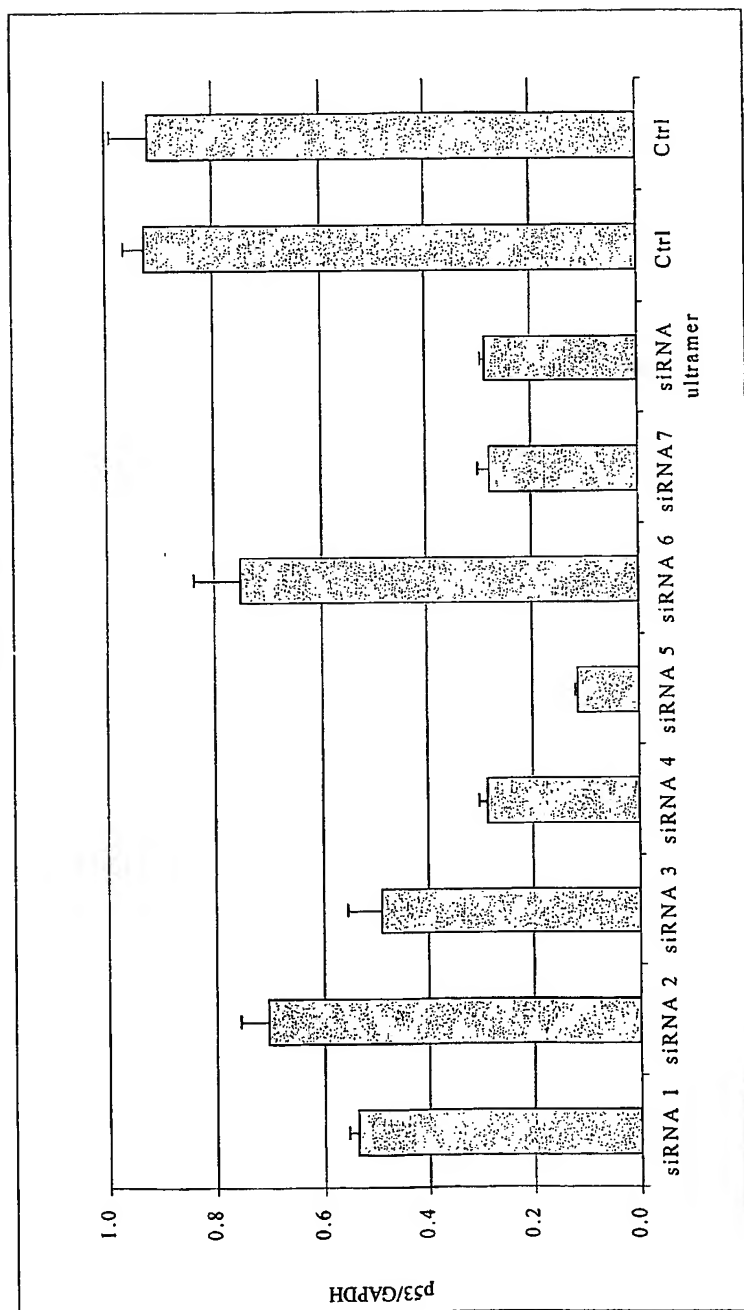


FIGURE 3.
GTP20 siRNA Efficacy
Expression of GTP20 normalized to GAPDH in Transfected hMSC

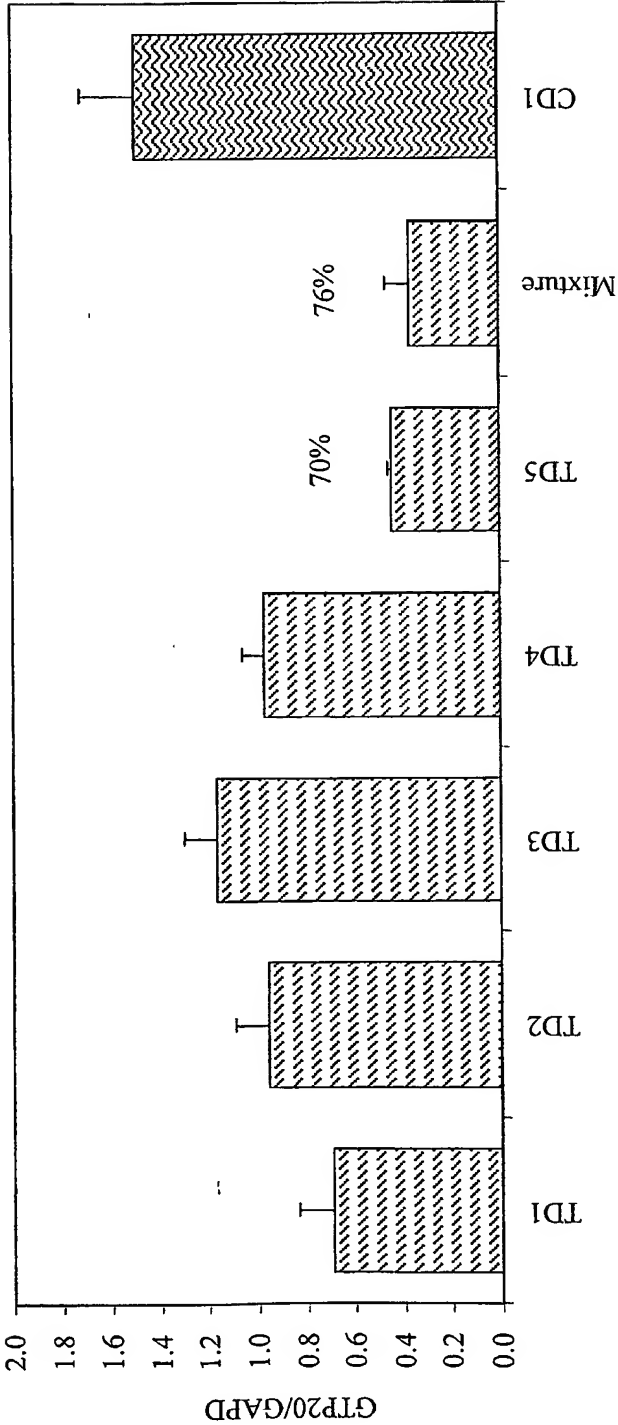
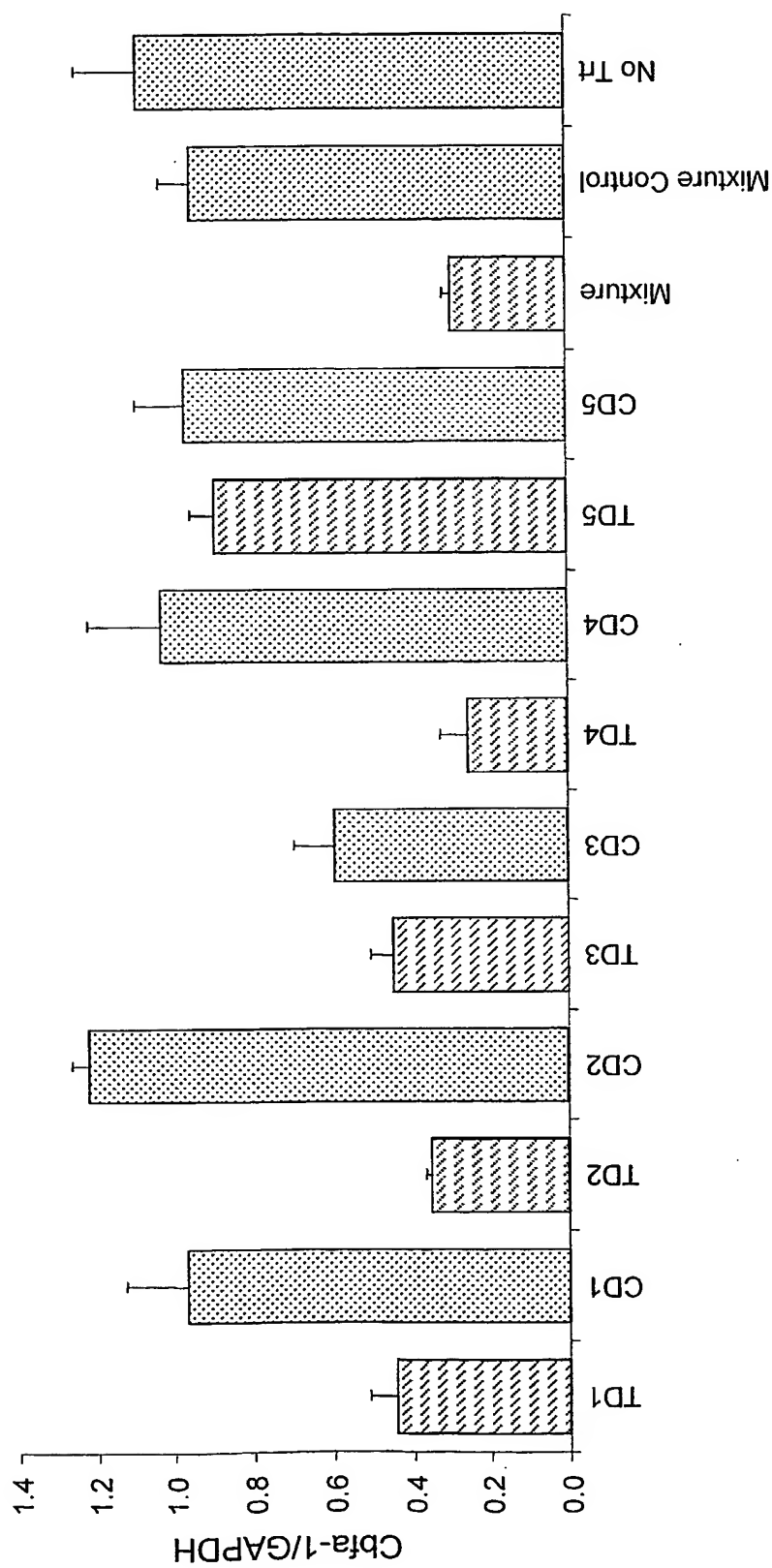


FIGURE 4.
Cbfa-1 siRNA Efficacy
 Expression of *Cbfa-1* normalized to GAPDH in Transfected hMSC



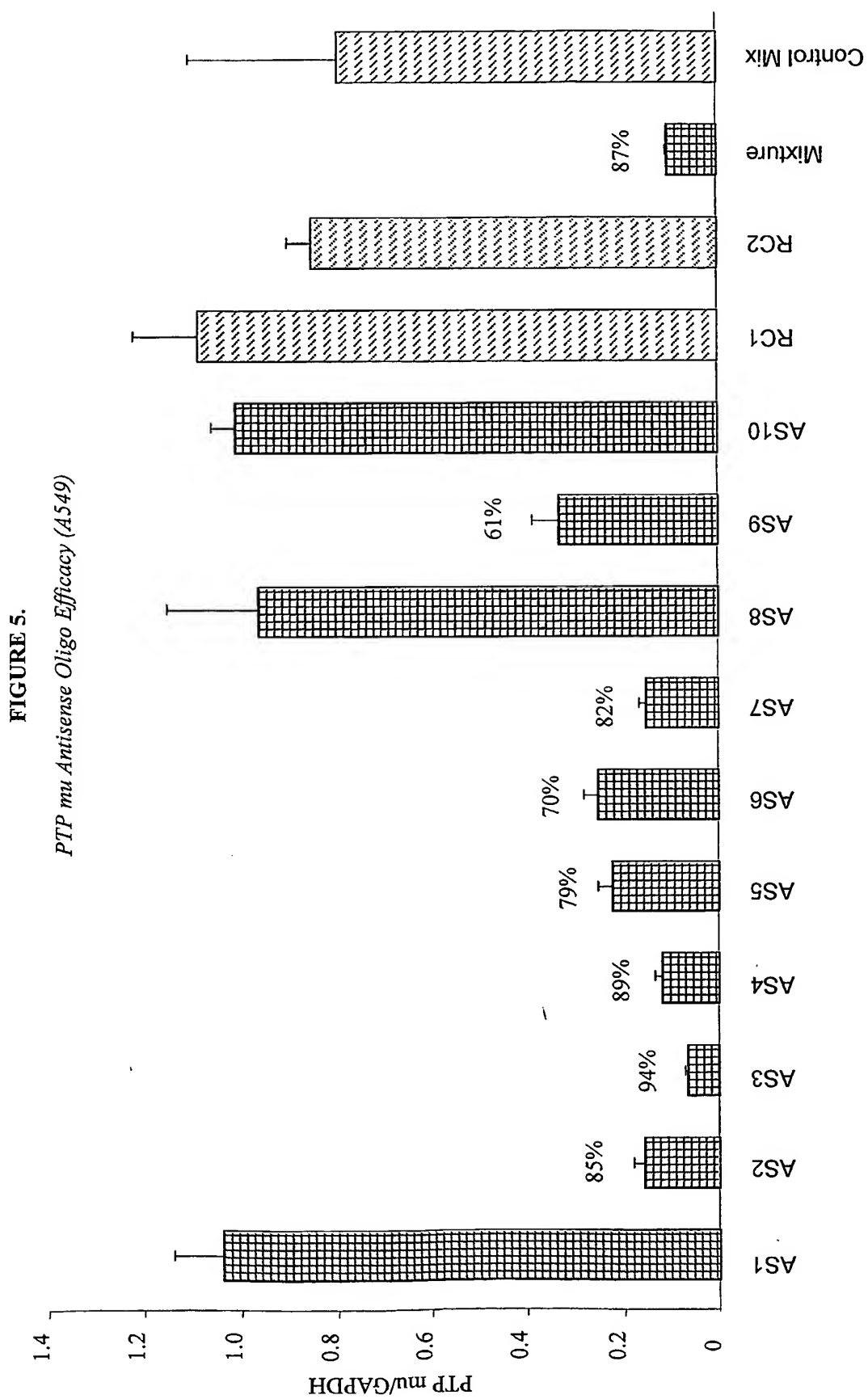


FIGURE 6.
PTP-PEST Antisense Oligo Efficacy (HuVEC)

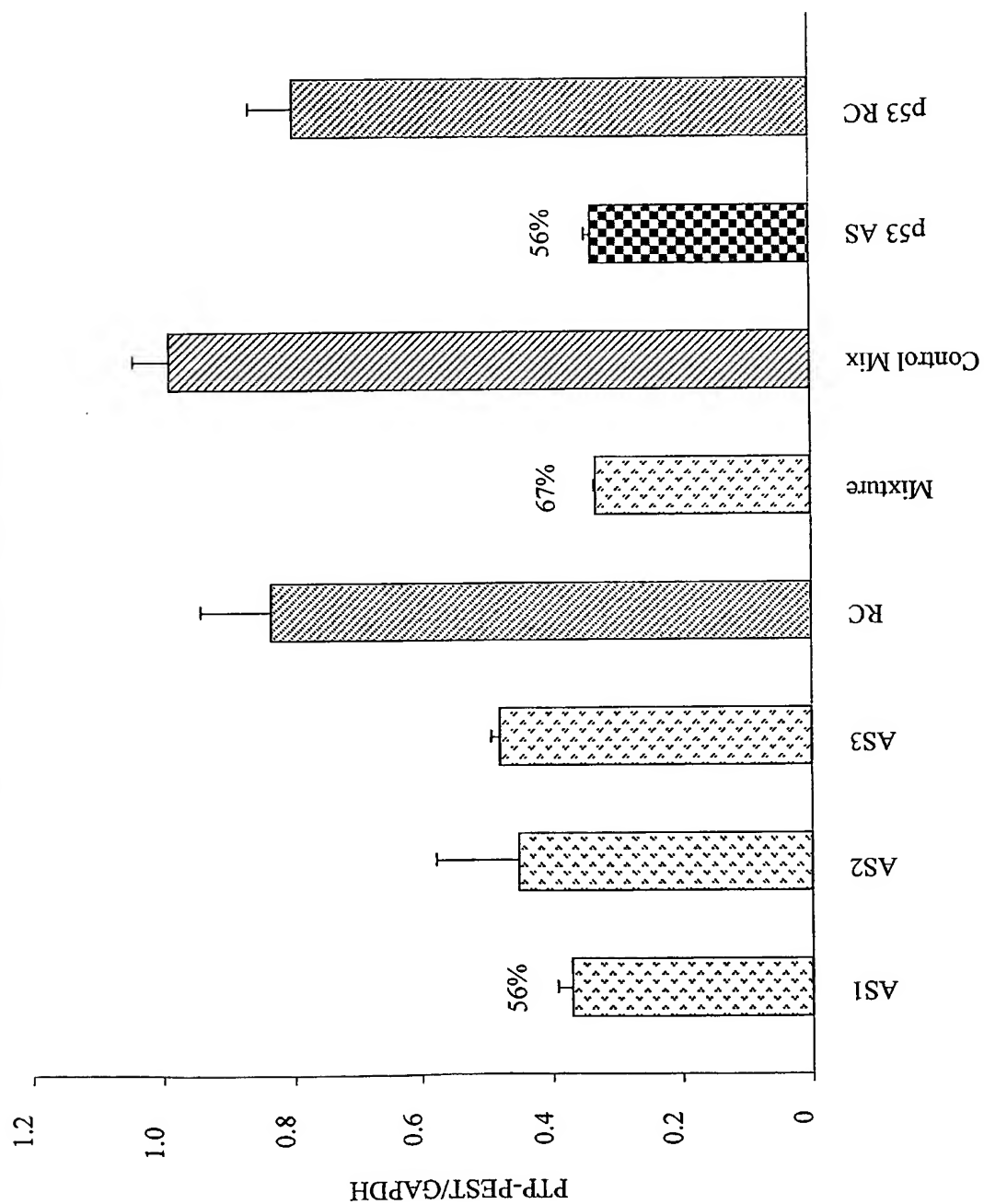
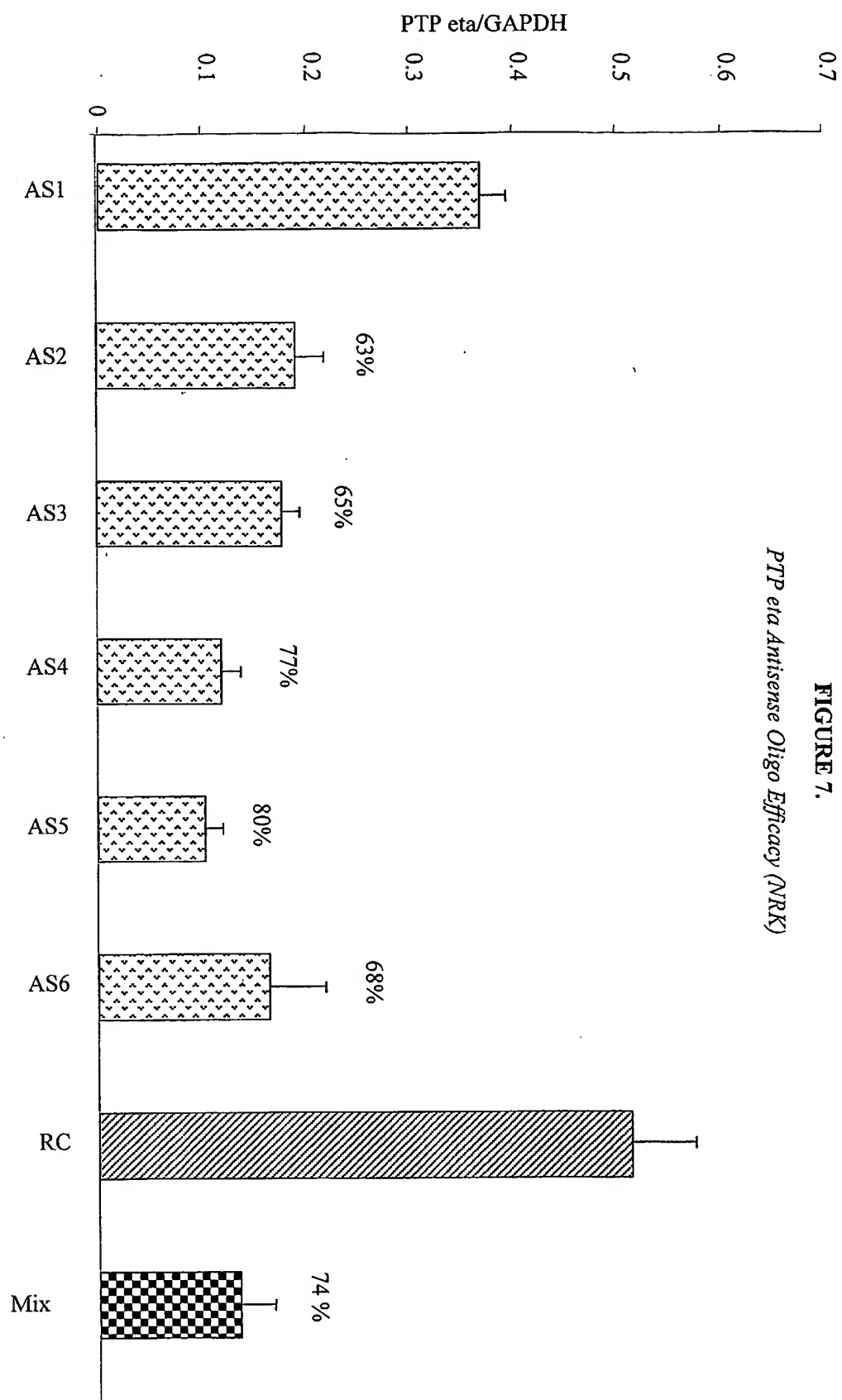


FIGURE 7.
PTP eta Antisense Oligo Efficacy (NRK)



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.